Exhibit H
Part 1

Getting Started with Conjoint Analysis

Strategies for Product Design and Pricing Research

Bryan K. Orme

Third Edition

Getting Started with Conjoint Analysis

Strategies for Product Design and Pricing Research

Third Edition

Bryan K. Orme

Research Publishers LLC

Getting Started with Conjoint Analysis:
Strategies for Product Design and Pricing Research
Third Edition
by Bryan K. Orme

Publisher
Research Publishers LLC
2711 N. Sepulveda Blvd. #320
Manhattan Beach, CA 90266-2725
USA
http://www.research-publishers.com

ISBN13 978-1-60147-111-6 (full text, paper)
ISBN13 978-1-60147-112-3 (full text, electronic, pdf)
ISBN13 978-1-60147-113-0 (full text, electronic, EPUB)
ISBN13 978-1-60147-114-7 (full text, electronic, Kindle)

Cover art by Paul Klee (1879–1940)
Maske Furcht, 1932, 286 (Y 6)
100 x 57 cm
Ölfarbe auf Jute auf Keilrahmen
The Museum of Modern Art, New York

Cover design by Jane Tenenbaum, Tenenbaum Design, Cambridge, Mass.
Electronic typesetting by Amy Hendrickson, TEXnology, Inc., Boston
Printed and bound by RR Donnelley, Breinigsville, Pa.

Printed in the United States of America. 9 8 7 6 5 4 3 2

Contents

Foreword vii

Preface ix

Acknowledgments xi

Figures xiii

Tables xv

Exhibits xvii

1 Managerial Overview of Conjoint Analysis 1

2 How Conjoint Analysis Works 7
- 2.1 Marketing Problem and Attribute List 7
- 2.2 Survey Design Plan 8
- 2.3 Credit Card Survey 9
- 2.4 Conjoint Analysis Utilities 11
- 2.5 Importance Scores 11
- 2.6 Conjoint Analysis as a Predictive Model of Choice 11

3 Understanding the Value of Conjoint Analysis 19
- 3.1 Realism Begets Better Data 19
- 3.2 Brand Equity 22
- 3.3 Strategic Pricing Research 24
- 3.4 Preference, Not Market Share 25

4 A Short History of Conjoint Analysis 29
4.1 Early Conjoint Analysis (1960s and 1970s) 30
4.2 Conjoint Analysis in the 1980s 32
4.3 Conjoint Analysis in the 1990s 34
4.4 Year 2000 and Beyond 35

5 Choosing a Conjoint Method 39
5.1 Traditional Full-Profile Conjoint Analysis 39
5.2 Adaptive Conjoint Analysis 42
5.3 Choice-Based Conjoint 45
5.4 Partial-Profile Choice-Based Conjoint 48
5.5 Adaptive Choice-Based Conjoint 49
5.6 Menu-Based Choice 50
5.7 Which Conjoint Method Should You Use? 51

6 Formulating Attributes and Levels in Conjoint Analysis 53
6.1 Present Appropriate Information 54
6.2 Follow Guidelines in Defining Attributes 54
6.3 Use Prohibitions Sparingly 57
6.4 Alternative-Specific Attributes 58

7 Sample Size Issues for Conjoint Analysis 61
7.1 Sampling Error versus Measurement Error 62
7.2 Binary Variables and Proportions 63
7.3 Continuous Variables and Means 64
7.4 Small Populations and the Finite Population Correction 65
7.5 Measurement Error in Conjoint Studies 66
7.6 Typical Sample Sizes and Practical Guidelines 69

8 Traditional Conjoint Analysis with Excel 71
8.1 Data Organization and Coding 73
8.2 Multiple Regression Analysis 77

9 Interpreting the Results of Conjoint Analysis 81
9.1 Nature of Quantitative Data 81

9.2 Conjoint Utilities 82
9.3 Counts 83
9.4 Attribute Importance 83
9.5 Sensitivity Analysis Using Market Simulations 85
9.6 Price Elasticity, Price Sensitivity, and Willingness to Pay 88

10 Market Simulators for Conjoint Analysis **93**
10.1 What Is a Market Simulation? 93
10.2 Applications of Conjoint Simulations 94
10.3 Introducing New Products 97
10.4 Estimating Demand Curves and Elasticities 99
10.5 Designing Products for Market Segments 102
10.6 Game Theory and Conjoint Analysis 106
10.7 Simulation Methods and Sample Sizes 108
10.8 Interpreting the Output of Market Simulators 108
10.9 Multi-Store Simulators 109

11 Maximum Difference Scaling **111**
11.1 Motivation for Maximum Difference Scaling 111
11.2 Efficient Data Collection Mechanism 113
11.3 Analysis of Maximum Difference Data 114
11.4 Maximum Difference Scaling Versus Conjoint Analysis 119
11.5 Concerns about Maximum Difference Scaling 120
11.6 Anchored Scaling for MaxDiff 121
11.7 Predictive Validity 122

12 Adaptive Choice-Based Conjoint **123**
12.1 Origins of Adaptive Choice-Based Conjoint 123
12.2 A New Approach to Data Collection 125
12.3 Adaptive Versus Standard Choice-Based Conjoint 130
12.4 Summary 132

13 Menu-Based Choice **133**
13.1 Bundling to Overcome Individual Reservation Prices 134

13.2 Menu-Based Choice as a Research Technique 135
13.3 Designing Menu-Based Choice Experiments 140
13.4 Analyzing Menu-Based Choice Experiments 141
13.5 Summary 142

14 How Conjoint Analysis Is Used in Industry 143

14.1 Application of Choice Models at Procter & Gamble 144
14.2 Using Conjoint Analysis to Facilitate Doctor-Patient Communication 145
14.3 Conjoint Analysis at General Motors Corporation 145
14.4 Conjoint Analysis within Boeing Employees Credit Union 147
14.5 ACA at the Canadian Department of Fisheries and Oceans 149
14.6 Using Discrete Choice Methods in Health Care and Education . . 149
14.7 Analysis of Women's Preferences for Place of Child Delivery in Rural Tanzania 151
14.8 Conjoint Analysis at Microsoft 152
14.9 Conjoint Analysis at Lifetime Products 154

Appendices

A Glossary 159

B Contributors 223

Bibliography 225

Index 231

Foreword

Comments on the first edition of *Getting Started with Conjoint Analysis*.

Bryan Orme's book on conjoint analysis promises to be a new and excellent addition to the literature. Currently, virtually every marketing research text provides at least a chapter or so on conjoint analysis. More ambitious academic works include the book by Gustafsson, Herrmann, and Huber (2000) and the monograph by Krieger, Green, and Wind (2005).

Orme's contribution is something else again. Rather than adopt a typical pedantic approach, Orme's book is a practical, no-nonsense guide to what happens when one designs, executes, and analyzes data from real, marketplace problems. It should appeal to academics and consultant-practitioners alike.

An essential feature of the book is its emphasis on specialized software that enables the analyst to implement real-world studies. (I suspect that Sawtooth Software's founder, Rich Johnson, has played a pivotal role in the software department.)

Over a thirty-five-year period, conjoint analysis has grown from a relatively crude approach to a method with a high degree of sophistication. High-profile applications have included the following:

- Design of AT&T's first cellular telephone
- Design and implementation of the EZ-Pass toll collection system
- Development of new varieties of Mama Celeste pizzas
- New logo design for the Baltimore Ravens football team
- U.S. Navy reenlistment benefits
- New services for the Ritz Carlton and Marriott hotel chains

The book is easy to follow, while at the same time being almost encyclopedic in its coverage of topics ranging from study design to the presentation of results to clients.

Different reviewers will probably have their own favorite chapters, a reflection of built-in biases. I particularly liked the chapter that deals with real cases provided by practitioners. This is a fine idea and provides incentives for consultants and company-based researchers to share their experiences regarding methodological (if not substantive) issues.

I also very much appreciated chapter 1, which explains in simple terms the main ideas of conjoint analysis and related techniques. Chapter 5 provides a very nice overview of conjoint method selection. I also liked the "short history of conjoint analysis" chapter describing how things got started and matured through the years.

Chapter 10 is a very important chapter in the book. This is where the application of conjoint simulators and optimizers comes into play. Most real-world studies depend on how the researcher posits the strategic and tactical questions that clients should get asked before any data collection gets underway.

In sum, I have found Sawtooth Software to be a one-stop-shopping experience for virtually any question dealing with trade-off analysis and related techniques. Orme's book should facilitate the learning process even more. Through Sawtooth Software's conferences and internal research on best practices, conjoint analysis has both matured and flowered as a practical set of tools for designing new (or refurbished) products and services.

As for me, the excellent glossary gets my vote for the best part of the book by far. This glossary must have taken eons to write. And the meticulous results justify the immense effort that has gone into the whole enterprise.

Paul E. Green, Professor Emeritus of Marketing
University of Pennsylvania
Spring 2005

Preface

Over the years, I have heard colleagues lament that there is not a single book that provides a good overview and introduction to conjoint analysis. Most conjoint books that are currently available tend to be academic and assume a solid background in statistics. The goal with this work is to offer a practical, accessible introduction to conjoint analysis appropriate for business managers involved in marketing and strategic planning, research analysts, and, of course, university students.

This work assembles and updates a series of introductory articles I have written over the years, previously published on my company's Web site:

http://www.sawtoothsoftware.com

Chapter 13 on Menu-Based Choice is new with this edition. Admittedly, this work is heavily steeped in the Sawtooth Software perspective. Yet, my intention is that readers who do not use Sawtooth Software's systems will find much of general applicability and value.

One of the greatest challenges in learning about conjoint analysis is grasping and reconciling the vocabulary. Not only is the terminology extensive, but different authors refer to precisely the same thing using different words. To help the reader, appendix A features a glossary of terms. If you encounter terms you do not recognize, you may want to check the glossary for assistance.

Bryan K. Orme, Sawtooth Software
Orem, Utah
Spring 2013

Acknowledgments

In assembling this work, I am indebted to Rich Johnson, who has served as a patient mentor to me over the last fifteen years and who has reviewed and helped edit these articles. Paul Green was immensely gracious to provide suggestions and write the foreword. It means the world to me that the "father of conjoint analysis" would do this. Sadly, we lost Paul during this last year. I am so grateful that he was such a gentleman and so gracious to write the foreword for our first edition of this work.

The chapter on industry applications of conjoint analysis would not have been possible without the generous contributions of Greg Rogers, Liana Fraenkel, Jim Christian, Calvin Bierley, Murray Rudd, Charles Cunningham, Margaret Kruk, Peter Rockers, Chris Chapman, and Bob Goodwin. Because these individuals do not work at consulting companies, they had little motivation other than the desire to contribute to the cause of championing tradeoff analysis and to help me personally. Some of these individuals had to secure the approval of their legal departments to get their stories released. Their persistence is deeply appreciated.

Many thanks also to Tom Eagle and Keith Chrzan for their review and comments on the glossary. These two are some of the finest people and most capable conjoint analysts I have known. They caught a number of mistakes. I accept responsibility for any remaining errors.

I am also grateful for my association with quality individuals who have contributed to my academic and professional development: Joel Huber, Dick Wittink, Jon Pinnell, Karlan Witt, Keith Chrzan, and Peter Zandan.

And, of course, this book would not have been possible without the vision and enthusiasm of Tom Miller at Research Publishers LLC. He and his able copy-editors, Gypsy Fleischman and Kristin Gill, transformed a document with rough graphics and prose into a truly professional piece. Tom went the extra mile in developing graphics and choosing a memorable cover to make sure this book had aesthetic appeal—a talent that escapes me.

My long-term employment at Sawtooth Software has been a dream experience. Thank you, Rich, for seeing my potential and bringing me to Sawtooth Software. I treasure my relationship with my Sawtooth colleagues and our customers as well. Chris King, former president of Sawtooth Software and a close friend, deserves many thanks for encouraging this effort.

Above all, I am indebted to two key women in my life: my mother Cecelia and especially my wife Chandra. Mom instilled in me a desire for education and excellence. My wife (who is ever so more intelligent than I) loves me despite my many flaws and works tirelessly in support of me and our seven children. Thank you, dear, for your sacrifice and devotion.

Bryan K. Orme, Sawtooth Software
Orem, Utah
Spring 2013

Figures

2.1 Line chart of part-worth utilities for attribute levels 13
2.2 Calculating importance scores 14
2.3 Bar chart of importance scores 15
3.1 Choice-based conjoint demand curves 23
3.2 Estimating brand equity using points of equal relative demand . . . 23
3.3 Estimating brand equity using points of equal price 24
5.1 Usage of Sawtooth Software conjoint methods 40
9.1 Results of sensitivity analysis 87
11.1 Combined importance scores by gender 116
A.1 Ideal point utility function 191
A.2 Ideal vector model 192
A.3 Part-worth utility function 209

Tables

5.1 Usage of Sawtooth Software conjoint methods 41
11.1 Best and worst percentages from counts 115
11.2 Scores from logit analysis (zero-centered scale) 117
11.3 Scores from logit analysis (zero-to-100 scale) 117
11.4 Scores from logit analysis (100-sum scale) 118

Exhibits

1.1 Conjoint rating question 3
1.2 Choice-based conjoint question 5
2.1 A conjoint card 8
2.2 Credit card design plan 9
2.3 Credit card survey 10
2.4 Estimating part-worth utilities from a conjoint survey 12
2.5 Final exercise 17
3.1 Importance survey questions 20
3.2 Pairwise trade-off question 21
3.3 Choice-based conjoint question 21
4.1 Conjoint card for automobiles 30
4.2 Johnson's trade-off matrix with rank-order data 31
4.3 A choice set for automobiles 33
5.1 ACA step 1: Rank attributes in terms of preference 43
5.2 ACA step 2: Rate importance of attributes 44
5.3 ACA step 3: Pairs using graded rating scale 44
5.4 Choice set for designer dress suits 45
6.1 Resolving prohibitions with the composite-attribute approach 58
6.2 Alternative-specific attributes 59
6.3 Alternative-specific choice task 60
8.1 Full-factorial experimental design 72
8.2 Excel spreadsheet with conjoint data 74
8.3 Excel spreadsheet with coded data 75
8.4 Modified data table for analysis with Excel 76
8.5 Conjoint analysis with multiple regression in Excel 78
9.1 Relative importance of attributes 84
10.1 Part-worth utilities across segments 103
10.2 Game theory strategies and outcomes 107
11.1 Typical ratings grid 112
11.2 MaxDiff question 113
12.1 ACBC build-your-own questions 126
12.2 ACBC screening questions 128

12.3 ACBC choice task . . . 129
13.1 Simple menu task . . . 136
13.2 Base model choice with configuration options . . . 137
13.3 Build-your-own product menu . . . 138
13.4 Fast-food bundling menu . . . 139
13.5 Drug prescription menu choice . . . 140
A.1 Choice among shopping experiences . . . 168
A.2 Choice task with constant alternatives . . . 173
A.3 Dual-response MaxDiff question format . . . 180
A.4 A full-profile concept . . . 186
A.5 Graded pairs scale for menu preferences . . . 188
A.6 Calculation of importance from part-worth utility . . . 193
A.7 MaxDiff question format . . . 199
A.8 Calculating mean squared error . . . 201
A.9 Choice-based conjoint example with *none* concept . . . 204
A.10 Data from a simple trade-off matrix . . . 219

Chapter 1

Managerial Overview of Conjoint Analysis

A great deal of market research commissioned today is descriptive in nature rather than predictive. Descriptive information is useful to characterize demographics, usage patterns, and attitudes of individuals. Beyond descriptive information, managers need survey research tools that can predict what consumers will buy when faced with the variety of brands available and changing product characteristics. It is precisely due to this focus that conjoint or trade-off analysis has become so popular over the last three decades.

Humans employ a variety of heuristics when evaluating product alternatives and choosing in the marketplace. Many products are made up of a dizzying array of features (e.g., laptops, cell phone calling programs, insurance policies, and manufacturing equipment), whereas some are more straightforward (e.g., yogurt, beverages, and light bulbs) and are mainly differentiated by brand, packaging, and price. How does the manager decide what product characteristics, packaging, and branding to use or what price to charge to maximize profits? And how does the consumer evaluate the offering vis-à-vis other alternatives in the marketplace?

To decide what product to sell, managers may use their own intuition or the recommendations of design engineers, or they may look to competitors for indications of what already works. These strategies are myopic and reactive. In consumer-oriented organizations, potential products are often evaluated through concept (market) tests. Buyers are shown a product concept and asked questions regarding their purchase interest, or new products are actually placed in test markets. These tests can be quite expensive and time consuming, and generally investigate just one or a few variations of a product concept. In some surveys, respondents are asked to rate brands and products or to check which brands and product features they prefer. None of these approaches by itself has been consistently successful and cost-efficient. Conjoint analysis uses the best elements of these techniques in a cost-effective survey research approach.

Back in the early 1970s, marketing academics (Green and Rao 1971) applied the notion of conjoint measurement, which had been proposed by mathematical

psychologists (Luce and Tukey 1964), to solve these complex problems. The general idea was that humans evaluate the overall desirability of a complex product or service based on a function of the value of its separate (yet conjoined) parts. In the simplest form, one might assume an additive model. Consider a PC purchase. A consumer browsing the Internet might see the following alternative:

Dell
3 GHz processor
2GB RAM
17-inch flat panel display
$899

Assuming that this consumer uses some internal, subconscious additive point system to evaluate the overall attractiveness of the offer, the unobserved scores (called part-worths) for the attributes of this product for a given buyer might be

Attribute	*Part-worth*
Dell	20
3 GHz processor	50
2 GB RAM	5
17-inch flat panel display	15
$899	30
Total utility	120

The estimated overall utility or desirability of this product alternative is equal to the sum of its parts, or 120 utiles. The trick is to obtain these scores from individuals for the variety of attributes we might include in the product or that our competitors might include. To do this reliably, one first develops a list of attributes and multiple levels or degrees within each:

Brand	*Processor*	*RAM*	*Display*	*Price*
Dell	2 GHz	2 GB	17-inch	$699
Acer	3 GHz	4 GB	19-inch	$899
HP	4 GHz	8 GB	21-inch	$1,099
Sony				

It is easy to see that there are many possible combinations of these attribute levels. In the 1970s, it became popular to print each of many product profiles on separate cards and ask respondents to evaluate them by ranking or rating. Consider the conjoint rating question in exhibit 1.1.

By systematically varying the features of the product and observing how respondents react to the resulting product profiles, one can statistically deduce (typically using linear regression) the scores (part-worths) for the separate features re-

How likely are you to purchase this computer?
Use a scale from 0 to 100, where 0 = *not at all likely*,
and 100 = *definitely would purchase*.

HP
3 GHz processor
4 GB RAM
17-inch display
$1,099

Your likelihood to purchase:

Exhibit 1.1. Conjoint rating question

spondents may have been subconsciously using to evaluate products. In contrast to answering direct questions about individual product features, conjoint survey respondents cannot simply say that all features are important—they must trade off different aspects of the product (as in real life), weighing alternatives that have both highly desirable and less desirable qualities.

Using the attribute list developed earlier, there are $(4 \times 3 \times 3 \times 3 \times 3)$ or 324 possible product profiles that could be considered. But what makes conjoint analysis work so nicely is that an individual respondent does not have to evaluate all possible product profiles. If we are willing to assume a simple additive model (which tends to work well in practice), each respondent needs to evaluate only a fraction of the total combinations. With our example, only about eighteen to twenty-four carefully chosen product concepts (using experimental design principles of independence and balance) would need to be evaluated to lead to a complete set of part-worth scores for each respondent for all sixteen attribute levels. The part-worth scores are useful for determining which levels are preferred, and the relative importance of each attribute. Once we know these scores, we can simply sum them to predict how each respondent would react to any of the 324 possible product profiles.

Although the scores on the attribute levels provide significant value in and of themselves, the real value of conjoint analysis comes from the what-if market simulators that can easily be developed, often within spreadsheets. It follows that if, for each respondent, we can predict the overall desirability for all possible product profile combinations (given the set of attribute levels we measured), we can also predict how each respondent might choose if faced with a choice among two or more competing profiles. For example, we can simulate what percent of

the market would prefer each of four PCs (described using the different brands and performance characteristics we measured) if available for purchase. These predictions across a sample of respondents are referred to as shares of choice or shares of preference.

Holding competitive offerings constant, managers can systematically vary the features of their own product profile (such as pricing or performance attributes) and observe what percent of the market would prefer their product under each condition. With conjoint simulators, managers can estimate demand curves and substitution effects, answering questions like, "From which competitors do we take the most share if we increase the processor speed?" They can assess cannibalization effects: "What happens to our overall share if we come out with another product with lesser performance at a lower price?" In essence, managers have the ability to estimate the results of millions of possible concept/market tests based on data collected in a single survey research project among, typically, 300 to 600 respondents. If additional information is included, such as feature costs, computer search algorithms can find optimal product configurations (holding a set of competitors constant) to maximize share, revenue, or profit.

Since the 1970s, as one might expect, additional improvements and refinements have been made to conjoint analysis. In the 1980s, a computerized version of conjoint analysis called Adaptive Conjoint Analysis (ACA) was developed, which could customize the conjoint interview for each respondent, focusing on the attributes, levels, and trade-offs that were most relevant to each individual (Johnson 1987b). As a result, more attributes and levels could be studied effectively. In the 1990s, it became popular to ask respondents to simply choose among product profiles rather than rate each profile individually on a numeric scale. The feeling was that buyers in the real world do not actually score each alternative on a rating scale prior to choosing—they simply choose. With choice-based conjoint (CBC), respondents answer perhaps eight to twenty choice questions such as the one in exhibit 1.2.

Although each question takes longer to read (because there are multiple alternatives to consider), choice-based conjoint questions seem more realistic and can include a *none* choice that can be selected if none of the products would appeal to the survey respondent. Developments in computationally intensive statistical methods (hierarchical Bayes estimation) made it possible to estimate a complete set of part-worth scores on each attribute level for each respondent (Allenby, Arora, and Ginter 1995). The results were typically better than with ratings-based conjoint, and the resulting "what-if" market simulators became more accurate in predicting actual market choices.

Today, thousands of conjoint studies are conducted each year over the Internet, via hand-held and mobile technologies, by fax, using person-to-person interviews, or by mailed paper surveys. Leading organizations are saving a great deal of money on research and development costs, successfully using the results to design new products or line extensions, to reposition existing products, and to make more profitable pricing decisions. (See chapter 14 for examples).

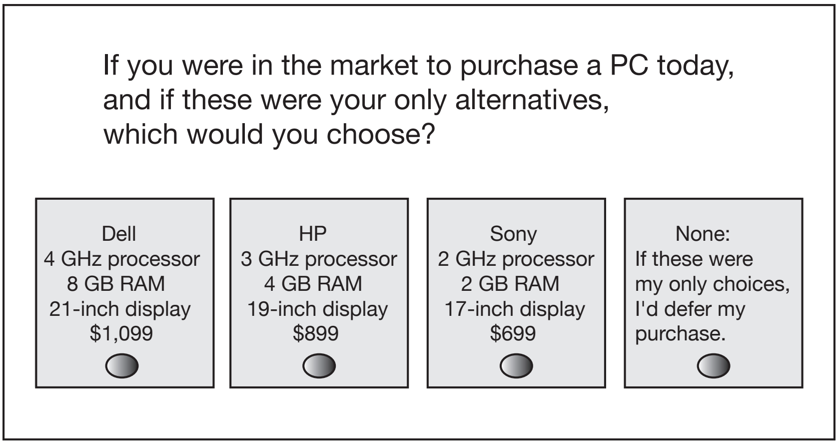


Exhibit 1.2. Choice-based conjoint question

Chapter 2

How Conjoint Analysis Works

Conjoint or trade-off analysis has become one of today's most commonly used market research tools for designing and pricing products and services. This chapter is designed to give newcomers to this exciting field some insight into how conjoint analysis works.

You will fill out a simple conjoint or trade-off survey dealing with different credit card offers. Please have a hand calculator available, so you can analyze the results of your survey. The calculations will be simple, involving only addition, division, and taking the average of three values at a time. I hope that by the end of this exercise I will have demystified conjoint analysis and you will have gained some insight into how this popular research method works.

2.1 Marketing Problem and Attribute List

Imagine that a credit card company is interested in how consumers trade off various aspects of credit card offerings, specifically, the brand, interest rate, and credit limit. A market researcher might approach the problem with conjoint analysis. First, a list of attributes and levels is developed that captures the range of brands, interest rates, and credit limits under consideration:

Brand	*Interest Rate*	*Credit Limit*
Visa	10%	\$2,500
MasterCard	15%	\$5,000
Discover	20%	\$7,500

With three attributes each at three levels, there are $(3 \times 3 \times 3) = 27$ possible credit card combinations. In the 1970s, researchers often printed each product profile in the conjoint survey on cards. Exhibit 2.1 is an illustration of one of the twenty-seven credit card combinations.

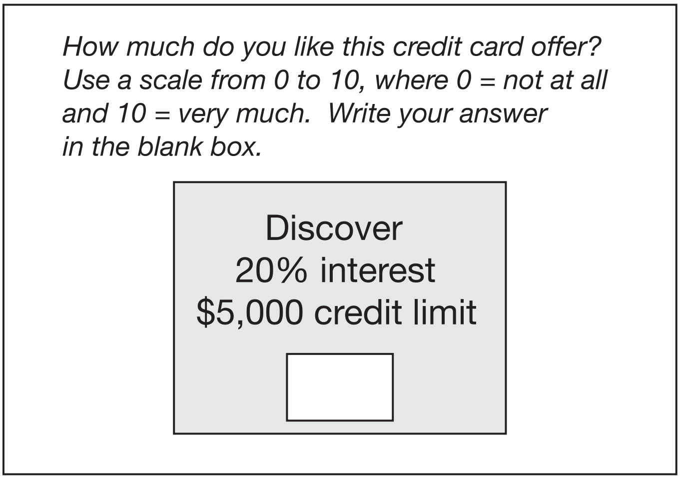


Exhibit 2.1. A conjoint card

With conjoint analysis, we rarely ask respondents to evaluate all possible combinations of attributes because this would create impossibly long questionnaires in most cases. Instead, we use an abbreviated survey design that focuses upon subsets of all possible combinations of attributes. Abbreviated survey designs may be found in reference books of experimental designs, or they may be generated by specialized computer programs.

2.2 Survey Design Plan

This chapter illustrates the simplest form of conjoint analysis: traditional full-profile conjoint. This is the original conjoint method (Green and Rao 1971). We show various credit card possibilities, referred to (coincidentally) as cards. The selected combinations of attributes or product profiles are called the design plan. Exhibit 2.2 shows the design plan for the conjoint analysis survey you are about to complete. Each row represents a conjoint analysis question or card. There are nine questions in the survey, with attribute combinations shown by the *X* marks in the exhibit. Exhibit 2.1 shows the first card of this survey: Discover, 20% interest, $5,000 credit limit.

This design plan has some valuable properties. First, you may notice that each level is shown exactly three times. Therefore, the plan is balanced. A clever aspect of this plan is that each level also appears exactly once with every other level from the other attributes. For example, Visa appears once at each of the levels of interest rate and credit limit. A similar relationship holds for the other two brands and for any combination of attributes taken two at a time. This property makes it possible to estimate the independent effect of each attribute with a relatively high degree

	Visa	MasterCard	Discover	10% interest	15% interest	20% interest	$2,500 credit limit	$5,000 credit limit	$7,500 credit limit
Card 1			X			X		X	
Card 2		X				X			X
Card 3	X				X				X
Card 4		X		X			X		
Card 5			X		X		X		
Card 6			X	X					X
Card 7	X			X				X	
Card 8		X			X			X	
Card 9	X					X	X		

Exhibit 2.2. Credit card design plan

of precision. For example, it would be difficult to distinguish the separate effects of brand and interest rate if Visa always appeared with the lowest interest rate. Would the preference for such a concept be due to the desirability of Visa or the low interest rate?

2.3 Credit Card Survey

Now that I have given an introduction to conjoint analysis and design plans, please complete a survey as if you were choosing a new credit card. Drawing upon the design plan, exhibit 2.3 presents nine credit card offers to be evaluated. For each credit card offer, you are asked, “How much do you like this credit card offer?” Use a ten-point scale to score each offer, where 0 means “not at all” and 10 means “very much.” After you have completed the credit card survey, rank the credit card offers in the final exercise (exhibit 2.5 on page 17).

Please complete the questions in exhibits 2.3 and 2.5 before reading further.

I hope you found the credit card survey interesting. You probably also found it a bit challenging. To evaluate the cards, you probably developed a strategy. Perhaps you decided early on which attribute was most important to you, and you based your decision mainly on that aspect. Perhaps each of the attributes carried about equal weight in your decision, and you needed to decide how much of one you were willing to give up for the other. You probably did not find an offer that was exactly what you wanted in every way. So it is with real purchase decisions. Consumers make these kinds of trade-offs every day.

Show how much you like each of the nine credit card offers below by writing your answers in the blank boxes.
Use a scale from 0 to 10, where 0 = not at all and 10 = very much.

1 Discover 20% interest $5,000 credit limit	2 MasterCard 20% interest $7,500 credit limit	3 Visa 15% interest $7,500 credit limit
4 MasterCard 10% interest $2,500 credit limit	5 Discover 15% interest $2,500 credit limit	6 Discover 10% interest $7,500 credit limit
7 Visa 10% interest $5,000 credit limit	8 MasterCard 15% interest $5,000 credit limit	9 Visa 20% interest $2,500 credit limit

Exhibit 2.3. Credit card survey

Now you are about to learn why conjoint analysis is so useful to managers and market research analysts. Based on the credit card ratings you provided, you will compute a weight or part-worth utility for each of the attribute levels. The set of weights would account for your overall credit card ratings. After all, you probably were loosely applying some sort of unconscious scoring mechanism or system of preference weights. Conjoint analysis seeks to uncover that system of preference weights. Human decision making is undoubtedly complex, and a simple set of weights can never fully capture the complexities. Conjoint analysis tends to do a very good job despite the model simplifications, providing more accurate insights and predictions of consumer behavior than most other research methods.

2.4 Conjoint Analysis Utilities

Because each attribute level appeared exactly once with every other level in the study, there is a simple way to estimate attribute level utilities (also known as part-worths). Of course, conjoint studies in the real world are rarely so straightforward. I have constructed this example so that the utility estimation may be done easily with a hand calculator. For this simple illustration, the utility estimate for each level is the average score for cards that include that level. Instructions for the calculations are provided in exhibit 2.4. Please compute the utility scores for each level now. If all has gone well, the utility scores should make intuitive sense to you. The higher the score for a level, the more desirable that level. It often helps to visualize data or to plot utility scores on a line chart. Please plot your part-worth utility values for each attribute in figure 2.1, drawing lines connecting the points on each graph.

2.5 Importance Scores

Some researchers calculate an importance score for each attribute. An importance score reflects the maximum effect each attribute has upon product choice, given the range of levels we included in the questionnaire. The calculations are straightforward, and you will again be able to use a hand calculator. Use figure 2.2 to work on your calculations. Refer to chapter 8 (page 84) or the glossary (page 193) for examples of importance score calculations. After you finish calculating importance scores, you can plot them on the bar chart in figure 2.3.

2.6 Conjoint Analysis as a Predictive Model of Choice

Charts of utilities and importance scores are useful, but a what-if market simulator that can be built using conjoint results is the most valuable tool for managers. A market simulator uses the utility scores to predict which product alternatives respondents would choose within competitive scenarios. The predictions can be made not only for the few product alternatives that were actually shown to respondents, but also for the often thousands or more potential combinations that were not shown.

Record your survey responses in the blank boxes for the numbered cards. The rating for each credit card offer is recorded in three groupings: once for the brand, for the interest rate, and for the credit limit.

Then compute the average rating across the three cards in each row.

Attribute Level	***Survey Responses (0-to-10 rating scale)***			***Row Average (Utiilty Score)***
Visa	Card 3	Card 7	Card 9	
MasterCard	Card 2	Card 4	Card 8	
Discover	Card 1	Card 5	Card 6	
10% interest	Card 4	Card 6	Card 7	
15% interest	Card 3	Card 5	Card 8	
20% interest	Card 1	Card 2	Card 9	
$2,500 credit limit	Card 4	Card 5	Card 9	
$5,000 credit limit	Card 1	Card 7	Card 8	
$7,500 credit limit	Card 2	Card 3	Card 6	

Exhibit 2.4. Estimating part-worth utilities from a conjoint survey

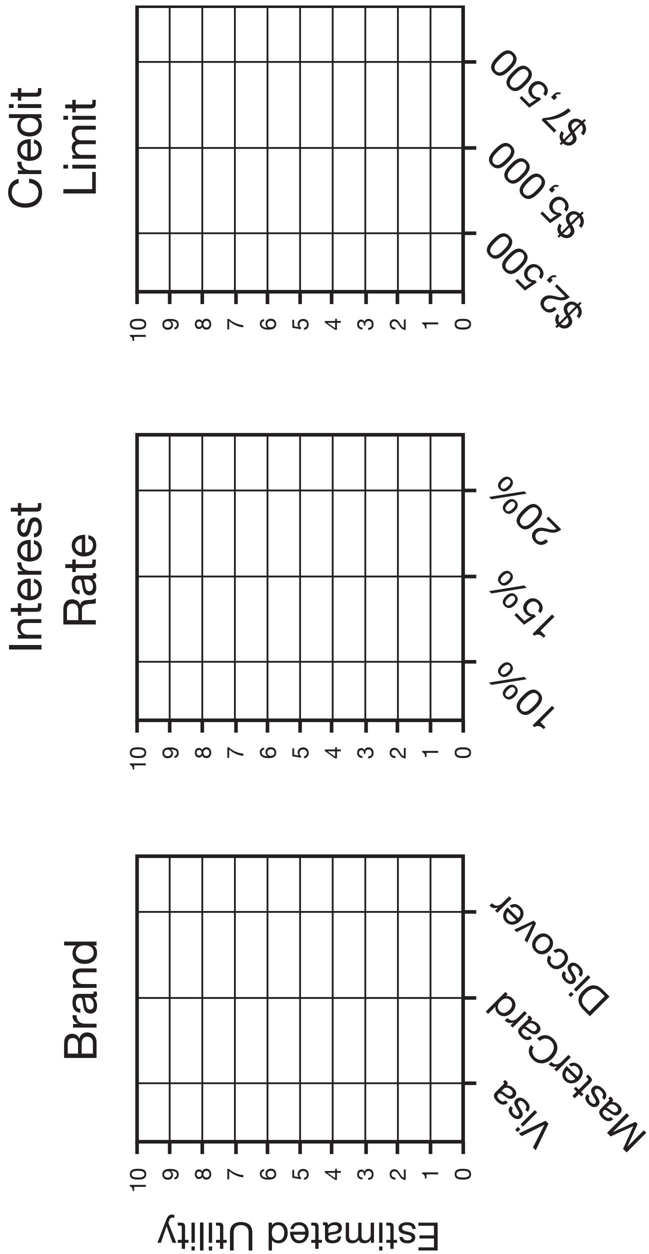


Figure 2.1. Line chart of part-worth utilities for attribute levels

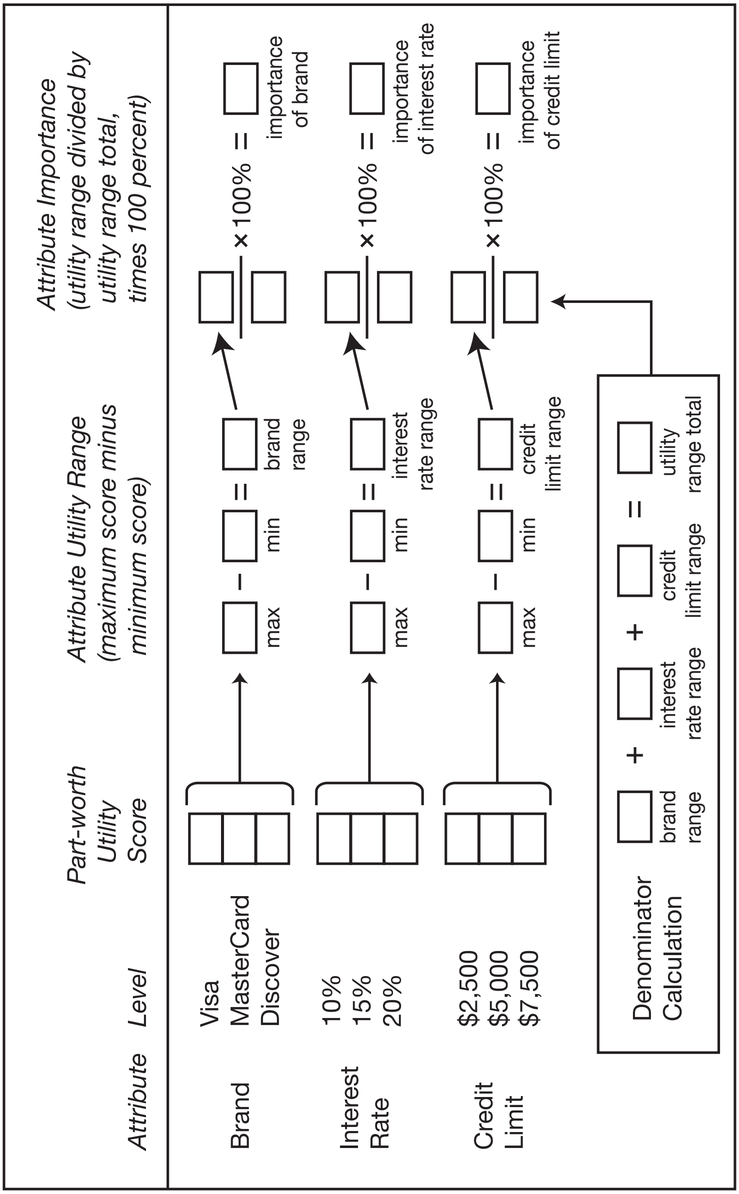


Figure 2.2. Calculating importance scores

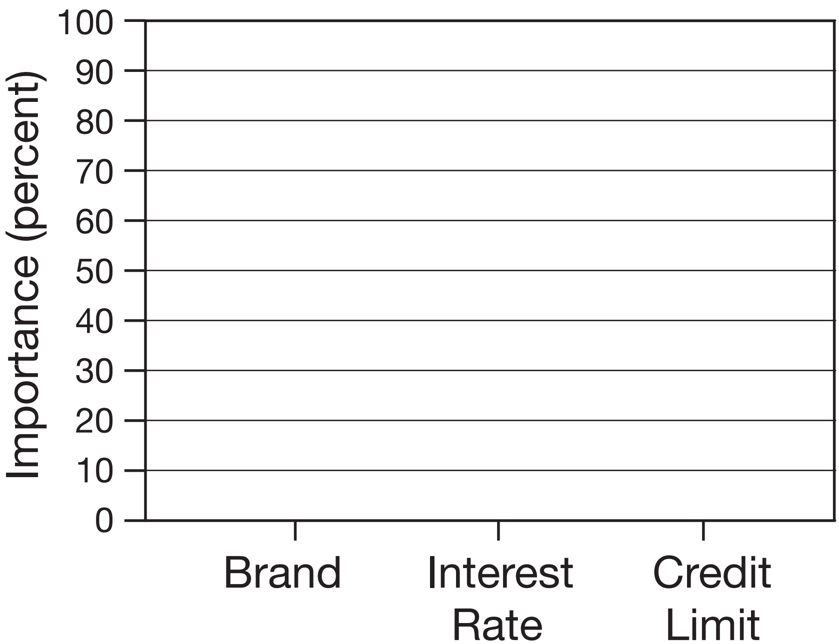


Figure 2.3. Bar chart of importance scores

In the example you have completed, there were only twenty-seven possible credit card configurations. You evaluated nine of them, from which you estimated part-worth utilities. But you also rank-ordered the three additional credit card offers in exhibit 2.5. These were holdout cards, credit card offers that were not used in developing your utility values. How well do your conjoint utility values predict how you ranked these three additional alternatives?

Add utility values that correspond to the attribute levels on each credit card to determine the total utility of that card (you need to add three part-worth values for each card). Compare the predicted total utilities to the rankings you previously provided in exhibit 2.5. If you completed the exercise carefully, the predictions should closely resemble your actual choices.

I hope this exercise has been useful and that you now understand how conjoint analysis works. If you answered the original and holdout questions carefully, it may surprise you how accurate the results from such a simple exercise can be. Of course, conjoint analysis in practice is more involved than what I have demonstrated here. Design plans are rarely as simple as portrayed here, and there are more powerful and accurate estimation techniques than taking simple averages. Chapter 8 expands upon these ideas, providing a brief introduction to estimating conjoint utilities using multiple regression. For more information about how to interpret your utilities and importance scores, see chapter 9 entitled "Interpreting the Results of Conjoint Analysis."

Discover 15% interest $7,500 credit limit	MasterCard 20% interest $5,000 credit limit	Visa 10% interest $2,500 credit limit
A	**B**	**C**

Considering the three credit card offers above, which is the best, the second best, and the worst?

Write letters in the spaces below:

☐ *is the best.*
A, B, or C

☐ *is the second best.*
A, B, or C

☐ *is the worst.*
A, B, or C

Exhibit 2.5. Final exercise

Chapter 3

Understanding the Value of Conjoint Analysis

Market researchers face two main challenges as they provide market intelligence for managers: to meet managers' objectives with useful, valid results and to communicate those results effectively. Failure on either of these points is fatal. Conjoint analysis provides useful results that, when presented well, are easy for managers to embrace and understand. It is no wonder that conjoint analysis is the most rapidly growing and one of the most widely used market research techniques today. This chapter discusses the benefits of conjoint analysis and finishes by highlighting a dangerous pitfall to avoid when presenting market simulators.

3.1 Realism Begets Better Data

Even though conjoint analysis involves sophisticated survey design and statistical analysis, and more effort by respondents, simpler approaches can be unrealistic, even useless. Suppose we were conducting a study about laptop computers, and using a survey like the one in exhibit 3.1. Respondents can answer importance survey questions quickly. A recent research project recorded an average time per response of five seconds (Orme 2003). Most respondents answer with high ratings, while the bottom half of the scale is largely ignored. This results in sub-par data for statistical analysis: skewed distributions, with typically little differentiation between attributes. Such self-explicated importances reveal little about how to build a better laptop. How much battery life will buyers trade off for a given increase in processor speed? Further, stated importances often do not reflect true values. It may be socially desirable to say price is unimportant—after all, respondents do not want to appear cheap. Yet in real-world laptop purchases, price may become a critical factor.

This chapter is adapted from an article published in *Quirk's Market Research Review*, March 1996.

When purchasing a laptop computer, how important is . . .

(Circle one number per item)

	Not Important								*Very Important*
Brand	1	2	3	4	5	6	7	8	9
Battery life	1	2	3	4	5	6	7	8	9
Processor speed	1	2	3	4	5	6	7	8	9
Weight	1	2	3	4	5	6	7	8	9
Price	1	2	3	4	5	6	7	8	9

Exhibit 3.1. Importance survey questions

Even though it is much easier on respondents to ask them to complete a grid such as shown in exhibit 3.1, these importance questions are not very meaningful. Buyers cannot always get the best of everything in the real world. They must make difficult trade-offs and concessions. When survey respondents (just like buyers) are forced to make difficult trade-offs, we learn the true value of product alternatives. And rather than ask respondents to react to generic terms like "battery life," we ask them to react to specific, realistic product specifications. The results are both meaningful and managerially actionable.

Conjoint analysis aims for greater realism, grounds attributes in concrete descriptions, and results in better discrimination among attribute importances. Conjoint analysis creates a more appropriate context for research. Consider a pairwise trade-off question featuring laptop computers. See exhibit 3.2.

Of course, conjoint questions can also be asked one product profile at a time, as in a traditional card sort. The rationale behind pairwise comparisons is this: People can make finer distinctions when they directly compare objects. For example, if someone hands you a four-pound rock, takes it away, and then hands you a five-pound rock, chances are you will not be able to tell which is heavier. But if you hold one rock in each hand, you will have a much better chance of guessing which weighs more. Despite the probable benefits of pairwise comparisons, we conducted a research study and found virtually no difference in the results for one-profile versus pairwise traditional conjoint analysis (Orme and King 1998).

Another flavor of conjoint analysis offers even greater realism and extends the idea of side-by-side comparisons: choice-based conjoint (Louviere and Woodworth 1983; Sawtooth Software 1993). For a choice-based conjoint question about laptop computers, see exhibit 3.3.

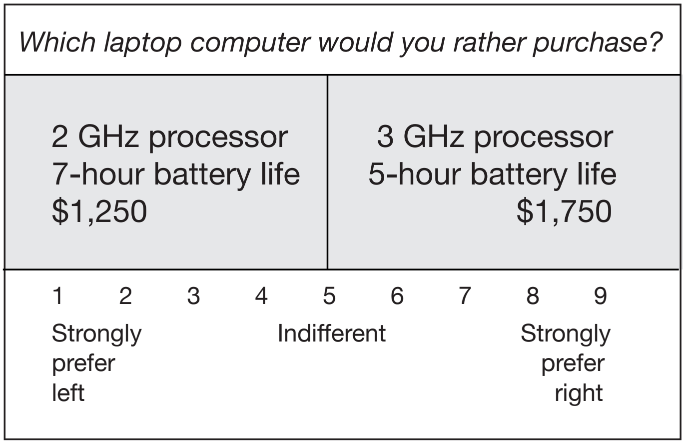


Exhibit 3.2. Pairwise trade-off question

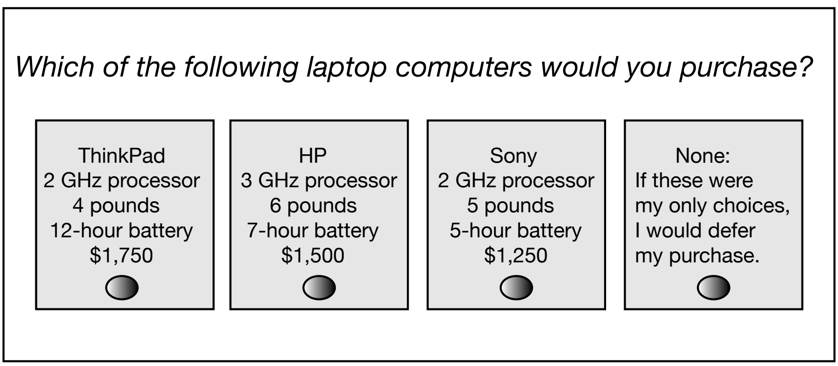


Exhibit 3.3. Choice-based conjoint question

Choice-based conjoint questions closely mimic what buyers do in the real world—choose among available offerings. Including *none* as an option enhances the realism, and allows those respondents who are not likely to purchase to express their disinterest. Choice-based data reflect choices, not just preferences that respondents have attempted to translate onto a ratings scale. If we agree that the ultimate goal of market simulators is to predict choice, then it is only natural that we would value choice-based data.

Some managers do not have the training in statistics to grasp the concept of orthogonal designs, main effects assumptions, or part-worth utility estimation. More technical folks, utilizing specialized software, can manage these details. Whether statisticians or otherwise, almost everyone can grasp the idea that realistic models should result from realistic questioning methods, and they can be comforted that conjoint analysis is a reliable, time-proven method.

3.2 Brand Equity

Conjoint analysis provides useful results for product development, pricing research, competitive positioning, and market segmentation. It can also measure brand equity, which is an especially critical issue for many managers.

Brand equity encompasses the intangible forces in the market that allow a product with a brand name to be worth more to buyers than one without. High-equity brands command higher prices and are less price elastic. Because brand equity goes directly to the bottom line, it is no surprise that managers are focused on it.

Choice-based conjoint offers a reliable way to measure brand equity. Choice-based conjoint presents respondents with varying product configurations and asks which they would purchase or choose. Each brand is presented at various prices throughout the interview. The percentage of times respondents choose each brand at each price point reveals preference and price sensitivity for the brands. Compelling demand curves result when we plot the probability of choice by price and connect the points with smooth lines. See figure 3.1 for hypothetical demand curves for three brands of pain reliever: Renew, Balmex, and PainFree.

If the brand manager for Renew wants to quantify the price premium it commands over the other brands, choice-based conjoint analysis reveals the answer. We can use the demand curves from figure 3.1 as a starting point: We draw a horizontal line through points A, B, and C representing a level of equal relative demand or preference. If Renew is priced at \$3.90 and Balmex at \$3.50, respondents on average will be indifferent (have the same preference) between the two. This forty-cent difference (point C price minus point B price or \$3.90 minus \$3.50) represents the premium or brand equity that Renew commands over Balmex. Similarly, Renew commands a sixty-cent premium over PainFree (point C price minus point A price). See figure 3.2.

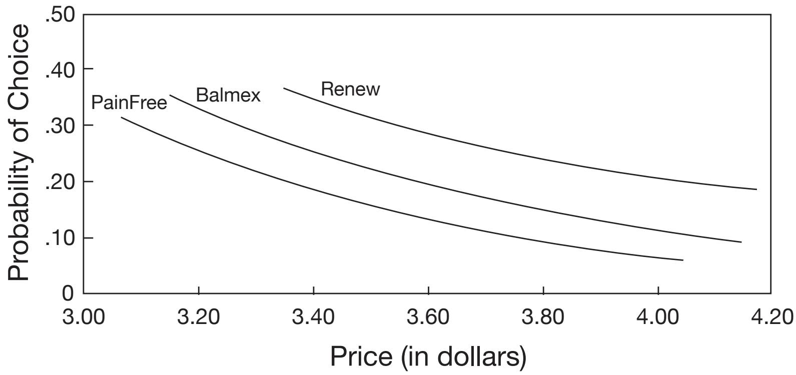


Figure 3.1. Choice-based conjoint demand curves

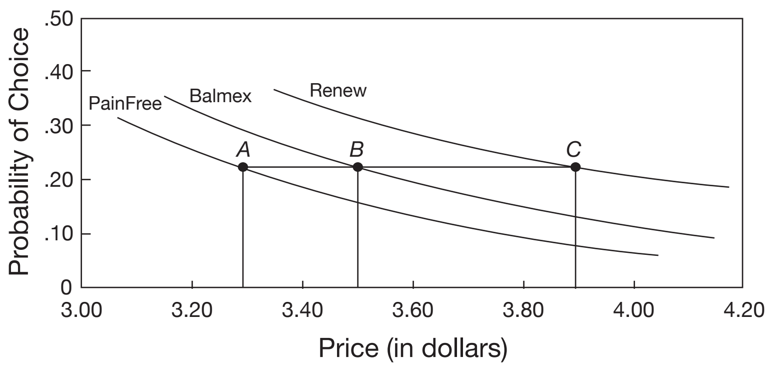


Figure 3.2. Estimating brand equity using points of equal relative demand

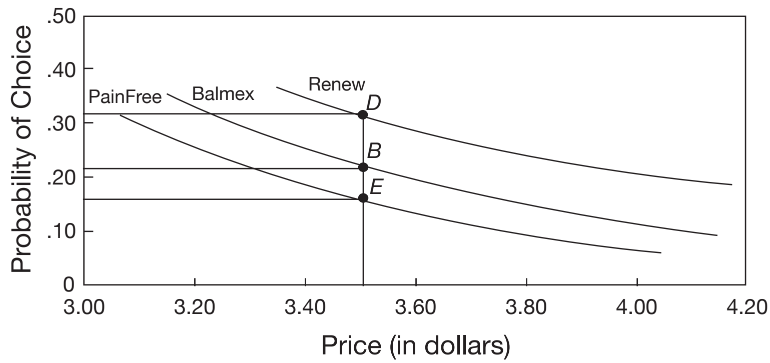


Figure 3.3. Estimating brand equity using points of equal price

Another approach to assessing brand equity results from comparing preferences with all brands offered at the same price. Imagine that we continue drawing the vertical line from \$3.50 through point B until it intersects Renew's demand curve. That point represents a relative preference or choice probability of 0.32. At \$3.50, Balmex and PainFree have choice probabilities of 0.22 and 0.16, respectively. See figure 3.3 with labeled points D, B, and E for Renew, Balmex, and PainFree, respectively, at the selected price point of \$3.50. Brand equity may be estimated by using ratios of choice probabilities or percentages. At the selected price point of \$3.50, Renew is preferred to Balmex by a ratio of $\frac{32}{22}$, or it has 45 percent higher preference than Balmex. Similarly, Renew is preferred to PainFree by a ratio of $\frac{32}{16}$ or 100 percent over PainFree.

3.3 Strategic Pricing Research

In an ideal world, researchers could accurately measure price sensitivity by manipulating prices in test markets and measuring changes in demand. While scanner technology and online shopping data have made this sort of analysis more feasible than ever before for many categories of consumer goods, these real-world experiments often face significant hurdles. Markets do not remain constant for the duration of the experiment. Macroeconomic forces can alter demand. Competitors can change their prices and/or promotions. Buyers can stock up to take advantage of lower prices. And new products may be introduced. While conjoint pricing experiments are not as realistic as the real-world event, conjoint experiments hold market forces constant. They can test price ranges or new products outside of current offerings.

In the demand curve example, Renew holds the enviable position of being preferred to Balmex and PainFree at all price levels. Notice also that the demand curves in exhibits 3.1 through 3.3 are not parallel. Renew's preference

declines at a slower rate than the other brands' as price increases. Respondents are less price sensitive toward Renew than the other brands. The ability to more directly measure unique price sensitivities by brand is an advantage choice-based conjoint enjoys over traditional main-effects-only conjoint analysis. While it is true that differential price sensitivities can be observed through sensitivity simulations from traditional full-profile conjoint analysis, most researchers believe that choice-based conjoint captures more accurate information about price sensitivity.

Demand curves provide strategic information for pricing decisions. Suppose Renew is the market leader. Renew's manager is considering initiating a price cut, and her past experience suggests that the discount brands will react with similar price cuts. She could learn a great deal using conjoint data—enough to avoid a mistake. The slopes of the demand curves show that, if prices were lowered, Renew would gain share at a slower rate than Balmex or PainFree. So if she lowers the price and the other brands follow, Renew's market share and profits would decrease.

Price elasticity can be quantified for each brand by examining the ratio of preference at the highest price versus preference at the lowest price. Alternatively, the price elasticity of demand (defined as percentage change in quantity demanded divided by percentage change in price) can be easily calculated for each brand in a choice-based conjoint study.

Some managers have been so pleased with this approach to strategic pricing research that they have funded wave after wave of conjoint tracking studies. They compare demand curves across time periods to quantify changes in brand equity, to gauge the results of previous pricing or other marketing mix changes, and to formulate future strategy.

Choice-based conjoint analysis has proven very useful and generally accurate for pricing decisions, especially when it comes to fast moving consumer goods. As an example, price sensitivity measurements by conjoint analysis for various Procter & Gamble products were shown to match well (on average) the price sensitivities calculated from econometric models applied to actual sales data (Renkin, Rogers, and Huber 2004).

3.4 Preference, Not Market Share

In the mid 1990s, we were involved in a choice-based conjoint study for a manufacturer of personal computers. Our main contact was the pricing manager whose objectives were to measure market awareness, preference, and price sensitivity for his sub-brands and major competitors. We conducted the study disk-by-mail and were soon delivering top-line conjoint results. This was prior to data collection over the Internet, so respondents received their computerized interviews on 3.5-inch floppy disks.

Our client was skeptical when he saw that the conjoint analysis reported that one of the company's newly released brands, call it FastPC, was preferred to its well-established brands. The client insisted that this could not be right and that

we check the data. We did—somewhat nervously, we might add—but found no errors. In the meantime, he called his sales department for a sanity check. Sales reported that FastPC was flying off the shelf. FastPC had exceeded all expectations.

While this happy-ending story warms us inside, it also illustrates a limitation of conjoint analysis. Conjoint analysis predicts preference, not market share. While the newly released FastPC was selling above expectations, its market share at that point fell short of established brands. Given enough time, adequate promotion, and distribution, we would expect FastPC's market share to align more closely with conjoint results.

Conjoint models do not predict market share due to a variety of reasons, including the following:

- Conjoint analysis assumes perfect information. In the conjoint interview, respondents are educated about available brands and features. In the real world, obscure brands have less chance of being purchased. Conjoint analysis cannot fully account for differences in awareness developed through advertising and promotion.
- Conjoint analysis assumes that all products are equally available. One brand is as conveniently selected as another in a conjoint interview.
- Respondents might not accurately reflect potential buyers. Many will not have the interest, authority, or ability to purchase.
- Results from conjoint analysis reflect the potential market acceptance of products and services, given proper promotion, distribution, and time.

Many researchers quantify factors that conjoint analysis cannot account for and build them back into the model using external effect adjustments. While this practice typically brings conjoint results more closely in line with actual market share, it draws us into a troublesome paradox. As more factors are accounted for and as we more accurately tune the conjoint model to market share, we start to believe that we have actually developed a valid market share predictor.

Believing that we have an accurate predictor of market share can lead us to misuse a model. That said, conjoint models are excellent directional indicators. Conjoint analysis can reveal product modifications that can increase market share, but it will probably not reveal how much actual market share will increase. Conjoint analysis can tell us that the market is more price sensitive for Brand A than Brand B, but we probably do not know the exact price sensitivity of either one. Conjoint analysis can identify which market segment will be most likely to purchase your client's product, but probably not the exact number of units that will be purchased.

The market simulator is usually the most anticipated deliverable for managers. Do not let this enthusiasm get out of hand. Conjoint simulators are directional indicators that can provide a great deal of information about relative feature importances and preferences for product configurations. While conjoint simulators are excellent tools for revealing strategic moves that can improve the success of a

product, they are not infallible market share predictors. Many other factors, such as awareness, distribution, advertising, and product life cycles, drive market share in the real world. Conjoint models can be fine-tuned to account partially for these elements, but we must avoid thinking that adjusted conjoint models can consistently and accurately predict volumetric absolutes such as market share. The only exception to this rule follows from careful validation based on real sales data, establishing a clear link between the calibrated conjoint model and sales volume for the specific product category and market in question.

Conjoint analysis increases the return on research dollars by providing managers with useful, valid information. Its realism leads to more accurate results and provides a strategic tool for quantifying brand equity and relative price sensitivity. To ensure success, researchers must carefully set management expectations regarding what conjoint analysis can and cannot do.

Chapter 4

A Short History of Conjoint Analysis

The genesis of new statistical models has rarely been within the field of marketing research. Marketing researchers have mainly borrowed from other fields. Conjoint analysis and the more recent discrete choice or choice-based conjoint methods are no exception. Conjoint methods were based on work in the sixties by mathematical psychologists and statisticians Luce and Tukey (1964), and discrete choice methods came from econometrics, building upon the work of McFadden (1974), 2000 Nobel Prize winner in economics.

Marketers sometimes have thought (or been taught) that the word "conjoint" refers to respondents evaluating features of products or services "CONsidered JOINTly." In reality, the adjective "conjoint" derives from the verb "to conjoin," meaning "joined together." The key characteristic of conjoint analysis is that respondents evaluate product profiles composed of multiple conjoined elements (attributes or features). Based on how respondents evaluate the combined elements (the product concepts), we deduce the preference scores that they might have assigned to individual components of the product that would have resulted in those overall evaluations. Essentially, it is a back-door, decompositional approach to estimating people's preferences for features rather than an explicit, compositional approach of simply asking respondents to rate the various features. The fundamental premise is that people cannot reliably express how they weight separate features of the product, but we can tease these out using the more realistic approach of asking for evaluations of product concepts through conjoint analysis.

Let us not deceive ourselves. Human decision making and the formation of preferences is complex, capricious, and ephemeral. Traditional conjoint analysis makes some heroic assumptions, including the proposition that the value of a product is equal to the sum of the values of its parts (i.e., simple additivity), and that complex decision making can be explained using a limited number of dimensions. Despite the leaps of faith, conjoint analysis tends to work well in practice, and gives managers, engineers, and marketers the insight they need to reduce un-

This chapter is based upon an article first published in *Quirk's Market Research Review*, July/August 2004.

Made in Europe
Rear-wheel drive
Four-door
$18,000

Exhibit 4.1. Conjoint card for automobiles

certainty when facing important decisions. Conjoint analysis is not perfect, but we do not need it to be. With all its assumptions and imperfections, it still trumps other methods.

4.1 Early Conjoint Analysis (1960s and 1970s)

Just prior to 1970, marketing professor Paul Green recognized that Luce and Tukey's (1964) article on conjoint measurement, published in a non-marketing journal, might be applied to marketing problems: to understand how buyers made complex purchase decisions, to estimate preferences and importances for product features, and to predict buyer behavior. Green could not have envisioned the profound impact his work on full-profile card-sort conjoint analysis would eventually achieve when he and coauthor Rao published their historic article "Conjoint Measurement for Quantifying Judgmental Data" in the *Journal of Marketing Research (JMR)* (Green and Rao 1971).

With early full-profile conjoint analysis, researchers carefully constructed a deck of conjoint cards based on published catalogs of orthogonal design plans. Each card described a product profile, such as shown in exhibit 4.1 for automobiles.

Respondents evaluated each of perhaps eighteen separate cards and sorted them in order from best to worst. Based on the observed orderings, researchers could statistically deduce, for each individual, which attributes were most important and which levels were most preferred. The card-sort approach seemed to work quite well as long as the number of attributes studied did not become too large. And researchers soon found that better data could be obtained by asking respondents to rate each card (say, on a ten-point scale of desirability) and using

	Made in USA	Made in Europe	Made in Far East
Front-wheel drive	7	6	3
Rear-wheel drive	9	8	5
All-wheel drive	4	2	1

Exhibit 4.2. Johnson's trade-off matrix with rank-order data

ordinary least squares regression analysis to derive the respondent preferences. In 1975 Green and Wind published an article in *Harvard Business Review* on measuring consumer judgments for carpet cleaners, and business leaders soon took notice of this new method.

Also just prior to 1970, a practitioner named Richard Johnson at Market Facts was working independently to solve a difficult client problem involving a durable goods product and trade-offs among twenty-eight separate product features, each having about five different realizations or levels. The problem was much more complex than those being solved by Green and coauthors with full-profile card-sort conjoint analysis, and Johnson invented a clever method of pairwise trade-offs. His paper on trade-off matrices was published in *JMR* (Johnson 1974). Rather than asking respondents to evaluate all attributes at the same time in full profile, Johnson broke the problem down into focused trade-offs involving just two attributes at a time. Respondents were asked to rank-order the cells within each table in terms of preference for the conjoined levels.

In exhibit 4.2 we see a respondent who liked the all-wheel drive vehicle made in the Far East best and the rear-wheel drive vehicle made in the United States least. With Johnson's trade-off matrices, respondents would complete a number of these pairwise tables, covering all attributes in the study (but not all possible combinations of attributes). By observing the rank-ordered judgments across trade-off matrices, Johnson was able to estimate a set of preference scores and attribute importances across the entire list of attributes for each individual. Because the method only asked about two attributes at a time, a larger number of attributes could be studied than was generally thought prudent with full-profile conjoint methods.

Near the end of the 1970s, academics Paul Green and Seenu Srinivasan published an influential paper in the *Journal of Consumer Research* summarizing the use of conjoint analysis in industry, outlining new developments, and giving advice regarding best practices (Green and Srinivasan 1978).

4.2 Conjoint Analysis in the 1980s

By the early 1980s, conjoint analysis was gaining in popularity, at least among leading researchers and academics possessing considerable statistical knowledge and computer programming skills. When commercial software became available in 1985, the floodgates were opened. Based on Green's work with full-profile conjoint analysis, Steve Herman and Bretton-Clark Software released a software system for IBM personal computers.

Also in 1985, Johnson and his new company, Sawtooth Software, released a software system (also for the IBM personal computer) called Adaptive Conjoint Analysis (ACA). Over many years of working with trade-off matrices, Johnson had discovered that respondents had difficulty dealing with the numerous tables and in providing realistic answers. He discovered that he could program a computer to administer the survey and collect the data. The computer could adapt the survey to each individual in real time, asking only the most relevant trade-offs in an abbreviated, more user-friendly way that encouraged more realistic responses. Respondents seemed to enjoy taking computer surveys, and some even commented that taking an ACA survey was like playing a game of chess with the computer.

One of the most exciting aspects of these commercial conjoint analysis programs for traditional full-profile conjoint and ACA was the inclusion of what-if market simulators. Once the preferences of typically hundreds of respondents for an array of product features and levels had been captured, researchers or business managers could test the market acceptance of competitive products in a simulated competitive environment. One simply scored the various product offerings for each individual by summing the preference scores associated with each product alternative. Respondents were projected to choose the alternative with the highest preference score. The results reflected the percent of respondents in the sample that preferred each product alternative, which was called share of preference. Managers could make any number of slight modifications to their products and immediately test the likely market response by pressing a button. Under the proper conditions, these shares of preference were fairly predictive of actual market shares. The market simulator took esoteric preference scores (part-worth utilities) and converted them into something much more meaningful and actionable for managers (product shares).

Conjoint analysis quickly became the most broadly used and powerful survey-based technique for measuring and predicting consumer preference. Helping to fuel this interest was an influential case study published by Green and Wind (1989) regarding a successful application of conjoint analysis to help Marriott design its new Courtyard hotels. But the mainstreaming of conjoint analysis was not without its critics, who argued that making conjoint analysis available to the masses through user-friendly software was akin to "giving dynamite to babies."

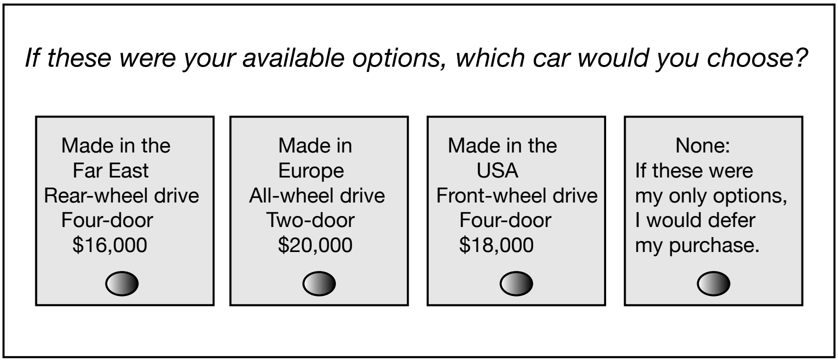


Exhibit 4.3. A choice set for automobiles

Those who experienced conjoint analysis in the late 1980s are familiar with the often acrimonious debates that ensued between two polarized camps: those advocating full-profile conjoint analysis and those in favor of ACA. In hindsight, the controversy had both positive and negative consequences. It certainly inspired research into the merits of various approaches. But it also dampened some of the enthusiasm and probably slowed the application of the technique. Some researchers and business managers paused to assess the fallout.

Prior to the release of the first two commercial conjoint analysis systems in 1985, Jordan Louviere and colleagues were adapting the idea of choice analysis among available alternatives and multinomial logit to, among other things, transportation and marketing problems. The groundwork for modeling choice among multiple alternatives had been laid by McFadden in the early 1970s. The concept of choice analysis was attractive: buyers did not rank or rate a series of products prior to purchase, they simply observed a set of available alternatives (again described in terms of conjoined features) and made a choice. From a theoretical and statistical standpoint, choice analysis was more defensible than ratings-based conjoint. But, from a practical standpoint, there were some challenges. A representative discrete choice question involving automobiles is shown in exhibit 4.3.

Discrete choice analysis seemed more realistic and natural for respondents. It offered powerful benefits, including the ability to do a better job of modeling interactions (i.e., brand-specific demand curves), availability effects, and cross-elasticities. Discrete choice analysis also had the flexibility to incorporate alternative-specific attributes and multiple constant alternatives. But the benefits came at considerable cost: discrete choice questions were an inefficient way to ask respondents questions. Respondents needed to read quite a bit of information before making a choice, and a choice only indicated which alternative was preferred rather than strength of preference.

With discrete choice there typically was not enough information to model each respondent's preferences. Rather, aggregate or summary models of preference were developed across groups of respondents. Aggregate models were subject to various problems such as independence from irrelevant alternatives (IIA or the red bus/blue bus problem) and ignorance of the separate preference functions for latent subgroups. Overcoming the problems of aggregation required building ever-more-complex models to account for attribute availability and cross-effects. These models, called mother logit models, were used by a relatively small and elite group of conjoint specialists throughout the 1980s. Given the lack of easy-to-use commercial software for fitting discrete choice models, most marketing researchers had neither the tools nor the stomach for building them.

4.3 Conjoint Analysis in the 1990s

Whereas the 1980s were characterized by a polarization of conjoint analysts into ideological camps, researchers in the 1990s came to recognize that no one conjoint method was the best approach for every problem, and expanded their repertoires. Sawtooth Software facilitated the discussion by publishing research from its users and hosting the Sawtooth Software Conference. User case studies demonstrated under what conditions various conjoint methods performed best. Sawtooth Software promoted the use of various conjoint methods by developing additional commercial software systems for full-profile conjoint analysis and discrete choice.

Based on industry usage studies conducted by leading academics (Vriens, Huber, and Wittink 1997), ACA was the most widely used conjoint technique and software system worldwide. By the end of the decade, ACA would yield that position to discrete choice analysis. Two main factors were responsible for discrete choice analysis overtaking ACA and other ratings-based conjoint methods by the turn of the century: (1) the release of commercial software for discrete choice modeling (CBC for choice-based conjoint) by Sawtooth Software in 1993 and (2) the application of hierarchical Bayes (HB) methods to estimate individual-level models from discrete choice data (principally due to articles and tutorials led by Greg Allenby of Ohio State University).

Discrete choice experiments are typically more difficult to design and analyze than traditional full-profile conjoint or ACA. Commercial software made it much easier to design and conduct CBC studies, while easy-to-use HB software made the analysis of choice data seem nearly as straightforward and familiar as the analysis of ratings-based conjoint. With individual-level models under HB, IIA and other problems due to aggregation were controlled or mostly solved. This has helped immensely with CBC studies, especially for those designed to investigate the incremental value of line extensions or me-too imitation products. While HB transformed the way discrete choice studies were analyzed, it also provided incremental benefits for traditional ratings-based conjoint methods. Traditional

conjoint methods had always estimated part-worth utilities at the individual level, but HB offered the prospect of more accurate estimation and shorter questionnaires. Other important developments during the 1990s included the following:

- Latent class models for segmenting respondents into relatively homogeneous groups, based on preferences
- Web-based data collection for all main flavors of conjoint and choice analysis
- Improvements in computer technology for presenting graphics
- Dramatic increases in computing speed and memory, making techniques such as HB feasible for common data sets
- Greater understanding of efficient conjoint and choice designs using concepts of level balance, level overlap, orthogonality, and utility balance
- Statistical Analysis System (SAS) routines for the design of discrete choice plans using computerized searches (Kuhfeld, Tobias, and Garratt 1994)
- Advances in the power and ease of use of market simulators offered both by commercial software developers and by consultants working with spreadsheet applications

The 1990s represented a decade of strong growth for conjoint analysis and its application in a fascinating variety of areas. Conjoint analysis had traditionally been applied to fast-moving consumer goods, technology products and electronics, durables (especially automotive), and a variety of service-based products such as cell phones, credit cards, and banking services. Other interesting areas of growth for conjoint analysis included design of Web sites, litigation and damages assessment, human resources and employee research, and Web-based sales agents for helping buyers search and make decisions about complex products and services. By the end of the decade, analysts had become so trusting of the technique that some used conjoint analysis to help them personally decide among cars to buy or members of the opposite sex to date.

4.4 Year 2000 and Beyond

Much recent research and development in conjoint analysis has focused on doing more with less: stretching the research dollar using IT-based initiatives, reducing the number of questions required of any one respondent with more efficient design plans and HB estimation, and reducing the complexity of conjoint questions using partial-profile designs.

Researchers have recently gone to great lengths to make conjoint analysis interviews more closely mimic reality: using animated three-dimensional renditions of product concepts rather than static two-dimensional graphics or pure text descriptions, and designing virtual shopping environments with realistic store aisles and shelves. In some cases the added expense of virtual reality has paid off in better data, in other cases it has not.

Since 2000, academics have been using HB-related methods to develop more complex models of consumer preference, relaxing the assumptions of additivity by incorporating noncompensatory effects, incorporating descriptive and motivational variables, modeling the interlinking web of multiple influencers and decision makers, and linking survey-based discrete choice data with sales data. Additional research includes efforts to customize discrete choice interviews so that they adapt to individual respondents in real time.

Interactive, customized discrete choice interviews can engage respondents in a dialog that probes their relevant decision space and reveals both compensatory (trade-off) and non-compensatory behavior (such as screening rules). It has long been held that buyers first screen available products to form consideration sets and then make choices within consideration sets. New research in adaptive CBC interviews has shown that staging the interview as a screening task (to select a consideration set) followed by focused trade-offs among considered products may lead to more accurate market simulation models, especially for high-involvement products and services described by many attributes (Gaskin, Evgeniou, Bailiff, and Hauser 2007; Johnson and Orme 2007).

Firms are becoming more nimble in the way they can customize products and services for consumers. Mass customization has become pervasive, as buyers can regularly design the product or service they wish to buy. Think of the Dell model for selling laptops a la carte, triple-play bundling of telecom services, restaurant menus, new car sales, banking and insurance options—these allow you the buyer to build your own product and then purchase it. The traditional conjoint approach does not mimic this buying process very well, so a new kind of conjoint called Menu-Based Conjoint or Menu-Based Choice (MBC) was created. The notion of MBC was envisioned and published at least a decade ago (Liechty, Ramaswamy, and Cohen 2001; Bakken and Bayer 2001). Only in the last few years have practitioners begun to more widely apply this newest conjoint-based methodology (Orme 2010; Orme 2012).

Software developers continue to make conjoint analysis more flexible, as well as faster and less expensive to carry out. Software systems often support multiple formats, including paper-based, PC-based, Web-based, and hand-held-device interviewing. Developers keep a watchful eye on the academic world for new ideas and methods that appear to be reliable and useful in practice. Commercially available market simulators offer more actionable information as they incorporate price and cost data, leading to market simulations of revenues and profitability rather than just shares of preference.

To reduce the amount of manual effort involved in specifying successive market simulations to find better products, automated search routines are now available. These find optimal or near-optimal solutions when dealing with millions of possible product configurations and dozens of competitors—usually within seconds or minutes. This has expanded opportunities for academics working in the area of game theory. These academics can study the evolution of markets as they achieve equilibrium, given a series of optimization moves by dueling competitors.

Importantly, more people are becoming proficient in conjoint analysis as the trade is being taught to new analysts. Academics are including more units on conjoint analysis in business school curricula. A growing number of seminars and conferences are promoting conjoint training and best practices. And research is being published and shared more readily over the Internet.

On the horizon, advances in the fields of neuromarketing and neuroeconomics seem particularly relevant to conjoint analysis. Rather than directly ask respondents to rate or choose among product concepts, the response to conjoint stimuli may be simultaneously measured on multiple dimensions using brain imaging technology. Rather than building a single model of part-worth utilities to predict choice, researchers might develop different utility functions related to the ability of product characteristics to "light up" different areas of the brain associated with (for example) euphoria, memories, risks, rational decision making, and fears. Such studies could help marketers gain insight into the key drivers operating within the psyche that lead respondents to choose what they do. While this area seems promising, imaging technology is currently expensive and time-consuming, and the interpretation of brain image scans involves many assumptions and uncertainties (Page and Raymond 2006; Green and Holbert 2012).

Yes, conjoint analysis is more than forty years old. But rather than stagnating in middle-age, it continues to evolve—transformed by new technology and methodologies, infused by new intellectual talent, and championed by business leaders. It is very much in the robust growth stage of its life cycle. In retrospect, very few would disagree that conjoint analysis represents one of the great success stories in quantitative marketing research.

Chapter 5

Choosing a Conjoint Method

It is paradoxical that many new developments in the conjoint analysis field have made the methods better than ever but have also made it more difficult to choose among methods. Limitations which earlier caused researchers to reject one flavor of conjoint analysis in favor of another have been overcome, thus blurring the lines of distinction between the approaches.

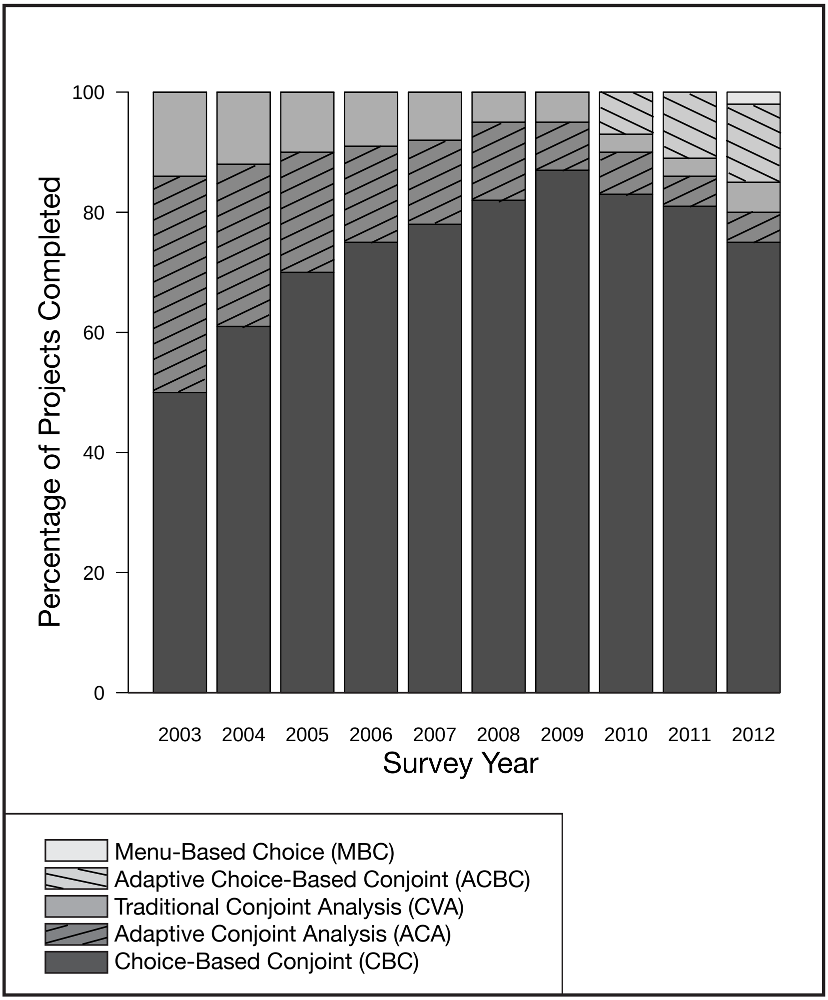
Conjoint analysis has become one of the most widely used quantitative tools in marketing research. When applied properly, it provides reliable and useful results. There are many conjoint methods. Just as the golfer doesn't rely on a single club, the conjoint researcher should weigh each research situation and pick the right combination of tools. It makes little sense to argue which method is the overall best approach. Each is designed to bring unique advantages to different research situations.

To get a feeling for the variety of conjoint analysis methods, consider the software packages offered by Sawtooth Software. Five systems are available: traditional full-profile conjoint analysis (also called conjoint value analysis or CVA), Adaptive Conjoint Analysis (ACA), choice-based conjoint (CBC), adaptive choice-based conjoint (ACBC), and menu-based choice (MBC). These five software offerings are representative of the main trade-off approaches used in industry today. According to a survey of Sawtooth Software customers, the relative use of these approaches in the Sawtooth Software community is CVA (5%), ACA (5%), and CBC (75%), ACBC (13%), and MBC (2%) (Sawtooth Software 2012). See figure 5.1 and table 5.1.

5.1 Traditional Full-Profile Conjoint Analysis

Sawtooth Software's version of traditional (ratings-based) full-profile conjoint analysis is called CVA. Other software firms, including SPSS Inc. and SAS Institute, also offer traditional conjoint analysis systems. Ratings-based conjoint was a

Adapted from a paper entitled "Which Conjoint Method Should I Use?" published in the Fall 1996 issue of *Sawtooth Solutions*. Interest in the paper, along with a steady flow of new developments in the conjoint analysis field, led us to update the piece many times since its original publication.



Source: Adapted from Sawtooth Software (2012).

Figure 5.1. Usage of Sawtooth Software conjoint methods

Table 5.1. Usage of Sawtooth Software conjoint methods

Survey Year	*Percentage of Projects Completed by Conjoint Method*				
	CVA	ACA	CBC	ACBC	MBC
2003	14	36	50	–	–
2004	12	27	61	–	–
2005	10	20	70	–	–
2006	9	16	75	–	–
2007	8	14	78	–	–
2008	5	13	82	–	–
2009	5	8	87	–	–
2010	3	7	83	7	–
2011	3	5	81	11	–
2012	5	5	75	13	2

CVA: Traditional Conjoint Analysis
ACA: Adaptive Conjoint Analysis
CBC: Choice-Based Conjoint
ACBC: Adaptive Choice-Based Conjoint
MBC: Menu-Based Choice

mainstay of the conjoint community for decades. Academics have suggested that the full-profile approach is useful for measuring up to about six attributes (Green and Srinivasan 1978). That number varies from project to project depending on the length of the attribute level text, the respondents' familiarity with the category, and whether attributes are shown as prototypes or pictures. Full-profile conjoint analysis may be used for paper-and-pencil studies, whereas ACA and adaptive choice-based conjoint (ACBC) must be administered via computer. Full-profile conjoint can also be used for computer-assisted personal interviews (CAPI) and Internet surveys.

Through the use of composite attributes, traditional full-profile conjoint can measure interactions between attributes such as brand and price. Composite attributes are created by including all combinations of levels from two or more attributes. For example, two attributes each with three levels can be combined into a single nine-level attribute. But interactions can only be measured in a limited sense with this approach. Interactions between attributes with more than two or three levels are better measured using CBC.

Sawtooth Software's CVA program can design pairwise conjoint questionnaires or single-concept card-sort designs. Showing one product at a time encourages respondents to evaluate products individually rather than in direct comparison with a competitive set of products. It focuses on probing the acceptability of a product offering rather than on the differences between competitive products.

If a comparative task is desired, CVA's pairwise approach may be used. Another alternative is to conduct a card-sort exercise. Though respondents view one product per card, in the process of evaluating the deck, they usually compare them side-by-side and in sets.

Because respondents see the products in full profile (all attributes at once), they tend to use simplification strategies if faced with too much information to process. Respondents may key in on a few salient attributes and ignore the others (Gilbride and Allenby 2004; Hauser, Dahan, Yee, and Orlin 2006; Johnson and Orme 2007). Huber (1997) points out that buyers in the real world may simplify tasks when facing complex decisions for certain categories, so simplification is not, by definition, always a bad thing.

Other researchers have suggested that different respondents simplify in different ways, and the summary effect across respondents should reflect the aggregate behavior of real buyers. As a counterpoint, a recent study we conducted at Sawtooth Software showed that respondents who took more time in a CBC study produced significantly different choices (on aggregate) than respondents who answered the questionnaire quickly (Johnson and Orme 2007). This would suggest that simplification behavior by survey respondents is not innocuous. It is more likely that survey respondents who simplify and answer questionnaires rapidly are answering in ways that, at both the individual level and in the aggregate, are not entirely consistent with how they would behave in real-world choices. Therefore, steps should be taken to encourage respondents to answer with more attention and effort in conjoint surveys, to the degree that this better approximates what they do in the real world.

5.2 Adaptive Conjoint Analysis

Released in 1985, Adaptive Conjoint Analysis (ACA) was Sawtooth Software's first conjoint analysis product. Other firms and academics have offered similar approaches, often referencing Sawtooth Software's product. ACA went on to become the most popular conjoint software tool and method in both Europe and the United States throughout the 1990s (Vriens, Huber, and Wittink 1997). Shortly after the turn of the century, CBC became more widely used—more on that later. ACA gained popularity early on because it was user-friendly for the analyst and respondent alike. But ACA's use has faded quite a bit over the last decade as choice-based methods, including ACBC, have proven more effective for most research situations.

ACA's main advantage was its ability to measure more attributes than was advisable with traditional full-profile conjoint or CBC. In ACA, respondents do not evaluate all attributes at the same time, which helps solve the problem of information overload that plagues many full-profile studies. Some academics and researchers have written that respondents cannot effectively process more than about six attributes at a time in a full-profile task. This is a useful guideline, but there is much debate about this. The number of attributes that respondents can

Rank these designers from most to least preferred.

Esprit
Liz Claiborne
Yves Saint Laurent

Exhibit 5.1. ACA step 1: Rank attributes in terms of preference

realistically manage in full profile depends on many factors, including the length of the attribute text, the use of graphics, and the respondents' sophistication and familiarity with the subject. ACA can include up to thirty attributes, although typical ACA projects involve eight to fifteen attributes. With six or fewer attributes, ACA's results are similar to the full-profile approach, though there is little compelling reason to use ACA in these situations.

In terms of limitations, the foremost is that ACA must be administered by computer. The interview adapts to respondents' answers as the survey progresses, which cannot be done via paper-and-pencil. Like most traditional conjoint approaches, ACA uses a main-effects model. This means that part-worth utilities for attributes are measured in an all-else-equal context, without the inclusion of attribute interactions. This can be limiting for studies in which it is important to estimate the price sensitivity for each brand. ACA has another limitation with respect to pricing studies: When price is included as one of the attributes, its importance is likely to be understated. The degree of understatement increases as the number of attributes studied increases. Recent studies have suggested that the importance of price can be understated by as much as two or three times, which is a significant problem (Pinnell 1994; Williams and Kilroy 2000).

ACA is a hybrid approach, combining stated evaluations of attributes and levels with conjoint pairwise comparisons. The first section of the interview employs a two-step self-explicated approach. Respondents rank or rate attribute levels (step 1), and then assign an importance to each attribute (step 2).

See exhibit 5.1 for an illustration of the ranking of attribute levels in terms of preference, which is step 1 of ACA. If Esprit were the most preferred designer and Liz Claiborne the least preferred, then the appropriate step 2 would be to ask the respondent for a rating of the importance of the difference between Esprit and Liz Claiborne, as illustrated in exhibit 5.2.

The self-explicated section of ACA puts emphasis on evaluating products in a systematic, feature-by-feature manner, rather than judging products as a whole or in a competitive context. The importance questions (step 2) are often chal-

If two designer dress suits were acceptable in all other ways, how important would this difference be?

Esprit versus Liz Claiborne

4 = extremely important
3 = very important
2 = somewhat important
1 = not important at all

Exhibit 5.2. ACA step 2: Rate importance of attributes

Which of these designer dress suits do you prefer?

Liz Claiborne Polyester $375	Yves Saint Laurent Silk $425

1	2	3	4	5	6	7	8	9
Strongly prefer left				*Indifferent*				*Strongly prefer right*

Exhibit 5.3. ACA step 3: Pairs using graded rating scale

lenging for respondents to answer reliably. Importance questions are the most problematic aspect of many self-explicated approaches, including step 2 of ACA surveys. Research presented at the Sawtooth Software conference suggested that dropping the importance questions from ACA surveys may result in better market share predictions and greater discrimination among attributes (as long as hierarchical Bayes is used to estimate the part-worths) (King, Hill, and Orme 2004). Subsequent analyses by Orme and Loke (2005) support these findings.

Using the information from the self-explicated section, ACA presents trade-off questions in the form of graded pairs. That is, two products are shown, and

Exhibit 5.4. Choice set for designer dress suits

respondents indicate which is preferred using a relative rating scale. See exhibit 5.3 for an illustration.

The product combinations are tailored to each respondent to ensure that each is relevant and meaningfully challenging. Each of the products is displayed in partial-profile, which means that only a subset (usually two or three) of the attributes is shown for any given question. Because of the self-explicated introductory section, the adaptive nature of the questionnaire, and the ratings-based conjoint trade-offs, ACA is able to stabilize estimates of respondents' preferences for more attributes using smaller sample sizes than CVA or CBC approaches.

Huber (1997) states that pairwise comparisons reflect the sort of purchase behavior wherein buyers compare products side-by-side. ACA does well for modeling high-involvement purchases, in which respondents focus on each of a number of product attributes before making a carefully considered decision. Purchases for low-involvement product categories described by only a few attributes along with pricing research studies are probably better handled using another method.

5.3 Choice-Based Conjoint

Choice-based conjoint analysis started to become popular in the early 1990s and from about the year 2000 on has become the most widely used conjoint technique in the world (Sawtooth Software 2012). CBC interviews closely mimic the purchase process for products in competitive contexts. Instead of rating or ranking product concepts, respondents are shown a set of products in full profile and asked to indicate which they would purchase. Exhibit 5.4 shows a choice set for designer dress suits.

As in the real world, respondents can decline to purchase in a CBC interview by choosing *none*. If the aim of conjoint research is to predict product or service choices, it seems natural to use data resulting from choices.

There are many variations of CBC questionnaires. Rather than asking respondents to choose one from each set of product concepts, some researchers ask

respondents to consider their next ten purchases, indicating how many of each product they would buy. This is known as chip allocation. Other researchers ask respondents to rank a full set of product alternatives or select the best and worst alternatives within each set. Most researchers favor the choice-based choose-one approach. It is not clear whether the additional effort of allocating choices or ranking (partial or complete) results in more accurate market-level models of buyer behavior, though it clearly increases the potential information content and respondent effort per completed task. The key question is whether taking that additional effort and instead applying it to answering more first-choice tasks is just as valuable, or maybe even more so.

Pinnell (1999) has suggested that, rather than using allocation in CBC tasks, researchers can first ask respondents what occasions lead to different choices, and then ask respondents to make discrete choices based on different (and customized) occasion scenarios. For example, with beer purchases, people consume different brands depending on the consumption occasion (at home alone or at a party with friends). With breakfast cereal purchases, the choice depends on the individual in the family that will eat the cereal. And the prescription of medications by doctors depends on the characteristics of the patient. For these examples, separate models could be built based on the specific occasion, with the results weighted across models to predict overall shares of choice.

Huber (1997) argues that choice tasks are more immediate and concrete than abstract rating or ranking tasks. They ask respondents how they would choose, given a set of potential offerings. Choice tasks show sets of products, and therefore mimic buying behavior in competitive contexts. Because choice-based questions show sets of products in full-profile, they encourage more respondent simplification than traditional full-profile questions. Comparing CBC to traditional full-profile conjoint or ACA generally shows that attributes that are important get greater emphasis, and attributes that are unimportant get less emphasis.

Sawtooth Software offers a popular software program for CBC. Choice tasks can be administered via computer-assisted personal interviewing, telephone interviewing, Internet surveys, or paper-and-pencil questionnaires. Another leading software provider, SAS Institute, provides superb routines for the design of CBC experiments. R and JMP are also common software platforms for the design of CBC experiments.

In contrast to either ACA or traditional full-profile conjoint (which automatically provide respondent-level part-worth preference scores), CBC results were originally analyzed at the aggregate or group level. But with the availability of latent class and hierarchical Bayes (HB) estimation methods starting in the late 1990s, both group- and individual-level analyses are accessible and practical. There are a number of ways to analyze choice results, the most popular of which are described below.

Aggregate Choice Analysis

Aggregate choice analysis was the first way to analyze data from choice-based conjoint studies. It was argued that aggregate analysis could permit estimation of subtle interaction effects (say, between brand and price) due to its ability to leverage a great deal of data across respondents. For most commercial applications, respondents cannot provide enough information with ratings- or sorting-based approaches to measure interactions at the individual level. While this advantage seems to favor aggregate analysis of choice data, academics and practitioners have argued that consumers have unique preferences and idiosyncrasies and that aggregate-level models that assume homogeneity cannot be as accurate as individual-level models.

Unless sophisticated cross-effects models are built, aggregate CBC analysis also suffers from its independence from irrelevant alternatives (IIA) assumption, often referred to as the red bus/blue bus problem. Very similar products in competitive scenarios can receive too much net share. Models that assume IIA fail when there are differential substitution effects across brands.

Latent Class Analysis

Latent class analysis addresses respondent heterogeneity in choice data and is often used to develop market segmentation (targeting) strategies. Instead of developing a single set of part-worths to represent all respondents (as in aggregate analysis), latent class simultaneously detects relatively homogeneous respondent segments and calculates part-worths for those discovered segments. If the market is truly segmented, latent class analysis can reveal much about market structure (including group membership for respondents) and improve predictability over aggregate choice models. Subtle interactions can also be modeled with latent class analysis. To summarize, latent class analysis has the benefits of aggregate estimation while recognizing market heterogeneity.

Hierarchical Bayes Estimation

Hierarchical Bayes estimation (HB) offers a powerful way of borrowing information from every respondent in the data set to improve the accuracy and stability of each individual's part-worth estimates. It has consistently proven successful in reducing the IIA problem and in improving the predictive ability of both individual-level models and market simulation share results. HB estimation can employ either main-effects-only models or models that also include interaction terms. But researchers are finding that many, if not most, of the interaction effects that were discovered using aggregate CBC analysis were actually due to unrecognized heterogeneity (Orme and Heft 1999). So main-effects models with HB are often sufficient to model choice. I will explain this further.

Suppose we have individual-level part-worths in a data set, and there are two types of respondents. One group prefers Honda and is less price-sensitive; the other prefers Kia and is more price-sensitive. If we perform sensitivity simula-

tions with no interaction terms included, we will see that the demand or market share of Kia changes more in response to price changes than the demand or market share of Honda. That is, respondents who initially prefer Kia are more likely to switch to Honda due to price changes than vice versa. Even with no interaction terms included in HB models, a brand/price interaction can be revealed through market simulations. HB models can reflect between-group differences in price sensitivity.

If interactions occur principally within individual preference structures (a person's disutility for spending depends on the brand), then explicitly modeling interaction terms may be necessary for accurate share predictions. Which approach is appropriate for your situation—models that represent respondent heterogeneity or models with interaction terms—may be difficult to tell. The benefits of individual-level part-worths make a compelling argument for HB estimation. I have consistently seen HB estimation outperform aggregate logit in predicting shares for holdout choices and shares in actual markets, even when there is only modest respondent heterogeneity.

There is some debate regarding which is better, latent class or HB, for analyzing conjoint analysis data. Rather than choose one or the other, the strengths of both can be leveraged during analysis. Latent class can be used for discovering meaningful segments of respondents, and HB can be used for developing individual-level part-worth utilities. The resulting preferences (estimated via HB) can be profiled by latent class segments.

5.4 Partial-Profile Choice-Based Conjoint

Researchers that favor choice-based conjoint over ratings-based approaches have looked for ways to increase the number of attributes that can be measured effectively using CBC. One solution that has gained mixed support over the last few years is partial-profile CBC (Chrzan and Elrod 1995; Chrzan 1999). With partial-profile CBC, each choice question includes a subset of the total number of attributes being studied. These attributes are randomly rotated into the tasks, so across all tasks in the survey each respondent typically considers all attributes and levels.

One problem with partial-profile CBC is that the data are spread quite thin. Each task has many attribute omissions, and the response is the less informative (though more natural) choice among alternatives. As a result, partial-profile CBC requires larger sample sizes to stabilize results relative to ACA. Despite this shortcoming, some researchers who used to use ACA for studying many attributes have shifted to partial-profile CBC. The individual-level parameter estimates have less stability than with ACA. But if the main goal is to achieve accurate market simulations, some researchers are willing to accept larger standard errors in individual-level estimates.

Partial-profile CBC results tend to reflect greater discrimination among most and least important attributes relative to ACA. Though with the latest versions of

ACA, one can probably remove this point of differentiation by omitting ACA's self-explicated importance questions. That is not to say that increased discrimination between attributes in terms of importance in conjoint is always the goal. But we do want to avoid artificial flattening due to self-explicated questions.

One obvious question that has not been resolved is whether partial-profile CBC is subject to the same price bias as ACA (i.e., understating price importance). We suspect that some of the price bias in ACA is due to the partial-profile nature of the process. Some early split-sample studies comparing partial- and full-profile CBC suggested that price bias was not a problem for partial-profile CBC. A number of more recent studies tend to point to lower price sensitivity estimates for partial-profile CBC compared to full-profile CBC.

Johnson, Huber, and Orme (2004) conducted a split-sample study that showed strong price bias for partial-profile choice. Another paper echoed this finding (Frazier and Jones 2004). Some researchers approach the problem of understating the importance of price by including price in each task and randomly selecting and ordering the remaining attributes. Including price in every choice task in partial-profile displays, however, does not necessarily resolve the concerns related to estimating price elasticity accurately using partial-profile choice. Recent studies suggest that price may still be understated even when included in each choice task.

5.5 Adaptive Choice-Based Conjoint

Newer CBC-related techniques combine the best aspects of adaptive interviewing with the realism and accuracy of choice data. Some researchers are first asking respondents about considered brands and features, and including just those within follow-up CBC exercises. Related to that, a new adaptive choice-based conjoint method (ACBC) developed by Sawtooth Software offers a question flow that incorporates the well-established theory that buyers make complex choices by forming a consideration set (typically using cut-off screening rules) and then choosing a product within that consideration set. ACBC displays relevant products for respondents to consider by patterning them after the preferred product that respondents have first specified using a build-your-own (BYO) exercise.

Another strength of ACBC is its ability to carry only a subset of relevant attributes and levels into the conjoint exercise. Attributes earlier deemed by the respondent to be of no consequence can be omitted from further consideration. Also, levels within attributes that are not in the respondent's consideration set can be omitted from further consideration.

Respondents find ACBC interviews more engaging, realistic, and relevant compared to traditional (static) CBC interviews. Even though the interviews are typically longer than standard CBC questionnaires, respondents generally prefer the overall experience. Sawtooth Software's ACBC questionnaire involves a progression of different-style choice exercises, which keeps things interesting. Plus, the products presented are centered around the respondent's preferred product

concept, so they are more realistic and relevant. A handful of comparisons have been made between ACBC and standard CBC, and the part-worth utility results are generally quite similar (but not identical). ACBC captures more information at the individual level, often leading to more accurate predictions than standard CBC, even given smaller sample sizes for ACBC. Plus, the improved accuracy of individual-level preferences leads to better discrimination between respondents and more stable detection of useful market segments than standard CBC when applying common techniques such as latent class or cluster analysis.

ACBC does not seem to be a replacement for standard CBC. For smaller and more standard problems involving four or fewer attributes (such as the common brand-package-price research), standard CBC approaches perform admirably. But for problems involving five or more attributes, for example, ACBC seems to be an excellent approach, as long as you can spare about eight to fifteen minutes of interview time. Sawtooth Software's ACBC is described more completely in chapter 12.

5.6 Menu-Based Choice

Some products and services are bought in a build-your-own (BYO) way. Common examples include buying laptop computers, restaurant meals, insurance policies, telecom services, new cars, and specialized industrial equipment. Such purchases often look like a menu in which certain features are bundled (typically for a discount) and others are offered a la carte. Buyers select the options they want, typically with prices attached to each component. And, after they are pleased with the final configuration and price, they make their purchase.

If you are studying a product category that is sold in a BYO way, then the conjoint approach of presenting a set of pre-configured products does not seem to capture faithfully the context of the buying process. Menu-based choice and CBC have been compared, and the two approaches capture fundamentally different preferences (part-worth utilities) and reflect different buying strategies (Johnson, Orme, and Pinnell 2006).

Although menu-based conjoint approaches have been described in the literature for at least a decade, their widespread use has only recently been seen. According to a 2012 survey of users of Sawtooth Software products, menu-based choice (MBC) studies accounted for about 2 percent of conjoint projects completed (Sawtooth Software 2012).

There are many approaches for designing MBC surveys and analyzing MBC data that work quite well. Sawtooth Software released its software tool for conducting MBC studies in 2012. MBC studies are more complex to design, program, and analyze than other types of conjoint surveys, so those who are relative newcomers to conjoint analysis may want to rely upon experienced consultants in the field.

5.7 Which Conjoint Method Should You Use?

You should choose a method that adequately reflects how buyers make decisions in the marketplace. This includes not only the competitive context but also the way in which products are described in words and displayed with multimedia or physical prototypes. It also includes how products may be considered by respondents. Do buyers build their preferred product (BYO) or choose among several pre-constructed alternatives? Is the product in a high-involvement category for which respondents deliberate carefully on all features, or is it in a low-involvement category for which respondents simplify the choice task and choose almost intuitively?

If you need to study many attributes, ACA or possibly partial-profile CBC have traditionally been considered. But, especially over the last ten years, the use of these approaches is waning. Partial-profile techniques are increasingly being viewed with suspicion, and researchers are generally displaying a greater amount of information in full profile within choice contexts. The traditional view that full-profile conjoint is limited to six attributes or fewer is being successfully challenged on a regular basis. Respondents seem to be able to deal with well-organized grids of information even better than early researchers and academics had supposed. Rather than try to limit respondents to comparing just two products at a time (to simplify the task), researchers are finding that respondents can often efficiently manage six or more product concepts at a time. And the resulting part-worth utilities tend to be better due to richer context. Displaying more rather than fewer product concepts per set often encourages respondents to reveal in-depth choice heuristics. In all cases, the number of attributes and product concepts we ask respondents to consider is a function of the amount of information that cleanly lays out on the page or computer screen.

In some areas of the world, survey populations do not have access to personal computers, and it may be too expensive to provide them. If your study must be administered by paper-and-pencil, this eliminates ACA or ACBC from consideration.

If you are dealing with small sample sizes (especially $n < 100$), you should be cautious about using CBC or MBC unless your attribute list is relatively concise and respondents are able to answer more than the usual number of choice questions. ACBC, ACA and traditional full-profile conjoint will stabilize estimates using smaller samples faster than CBC or MBC. See chapter 7 for a more complete discussion of sample size considerations.

For packaged goods research involving brand, package, and price, CBC with realistic store-shelf displays (often displaying dozens of product alternatives) is a robust approach. Increasingly, it is becoming hard to find reasons to prefer traditional ratings-based conjoint (CVA) or the ratings-based, partial-profile ACA method for general marketing research applications. CBC, as well as methods that leverage adaptive questioning and choice (such as ACBC), will dominate the conjoint landscape over the next decade.

If people buy the product or service by proactively configuring the product, then a menu-based conjoint approach like MBC is better suited for studying their preferences. MBC can study issues involving price sensitivity of items and bundling discounts. It can help firms create appropriate bundles of options that appeal to buyers. But MBC requires relatively large sample sizes—typically greater than 200 respondents, and often 600 or more.

Chapter 6

Formulating Attributes and Levels in Conjoint Analysis

Defining proper attributes and levels is arguably the most fundamental and critical aspect of designing a good conjoint study. It is also often the most time-consuming step in executing a conjoint analysis project. An attribute (sometimes called a factor) is a characteristic of a product (e.g., color), made up of various levels (there must be at least two for each attribute) or degrees of that characteristic (e.g., red, yellow, blue). The underlying theory of conjoint analysis holds that buyers view products as composed of various attributes and levels. Buyers place a certain part-worth utility on each of those characteristics, and we can determine the overall utility of any product by summing the value of its parts or levels.

In conjoint experiments, respondents express their preferences for products described by varying levels of attributes. By observing how respondents evaluate products in response to changes in attribute levels, we can estimate the impact each attribute level has upon overall product preference. That is, we can estimate utilities associated with attribute levels. After we learn respondents' preferences for the various attribute levels, we can predict how buyers might respond to a product with any potential combination of levels in our study, whether or not that actual product was displayed during the interview.

Typical full-profile conjoint studies in practice involve about eight or fewer attributes, each described on about two to five levels. Adaptive Conjoint Analysis (ACA) and adaptive choice-based conjoint (ACBC) studies often include many more attributes, each also described on about two to five levels. Some attributes are nominal (e.g., color, brand), and we cannot know ahead of time whether respondents prefer one level over another. Some attributes are ordinal (e.g., expedited delivery versus normal delivery), and we know ahead of time that rational respondents would usually prefer the levels in a certain order. Other attributes, such as weight, speed, amount of money, or length of time, are quantitative in nature, with the order of levels determined by the attributes and objects being evaluated.

6.1 Present Appropriate Information

Attribute descriptions should be concise statements with concrete meaning. Avoid using ranges to describe a single level of an attribute, such as "weighs 3 to 5 kilos." Rather than leaving the interpretation to the respondent, it would be better to specify "weighs 4 kilos." Levels such as "superior performance" also leave too much in question. What does "superior performance" mean? Try to use specific language to quantify (if possible) the exact meaning of the level.

Attributes that cannot be adequately described in words should be represented in multimedia. But if attributes do not require multimedia to adequately communicate their properties, it would probably be a mistake to make them multimedia. Though a multimedia interview might seem more attractive to respondents, it might bias the results in favor of multimedia attributes.

Present just the right amount of information, neither too little nor too much. Some respondents have a difficult time dealing with more than about six to eight attributes in full-profile conjoint methods like CBC. When faced with too much information, respondents often resort to simplification strategies to deal with the difficulty of the task (Green and Srinivasan 1978). Unless respondents employ the same sort of simplification strategies when making real-world decisions, full-profile results may place too much emphasis on the few most important features.

6.2 Follow Guidelines in Defining Attributes

Attribute definition is central to conjoint study design. Assembling the right combinations of attributes and attribute levels is critical to the success of conjoint studies. This section provides guidelines for defining attributes and attribute levels for conjoint research. The guidelines apply to all varieties of conjoint analysis.

Cover the Full Range of Possibilities for Attributes

Attribute levels should cover the full range of possibilities for relevant existing products as well as products that may not yet exist, but that you want to investigate. A market simulator allows you to extrapolate and interpolate. Interpolation is likely to produce acceptable results, but extrapolation is prone to error and should be avoided. One way to ensure that you are including the appropriate levels and ranges is to ask your client to specify ahead of time the market simulations to be run during the analysis phase of your study. That exercise can often reveal weaknesses in your attribute specifications.

Use Independent Attributes

Attributes should be independent. This is especially important for partial-profile conjoint studies such as ACA and partial-profile CBC. With partial-profile or hybrid conjoint (ACA involves both), attributes that overlap in meaning can get "double counted," resulting in too much inferred influence on product choice. It is therefore important to economize; including attributes with overlapping meanings is wasteful.

Furthermore, levels for related attributes may not combine naturally with one another. Though it can lead to more realistic interviews, it is often detrimental and sometimes fatal to prohibit levels from occurring with others.

Define Mutually Exclusive Attribute Levels

Levels within each attribute should be mutually exclusive. This point becomes clear when you specify products using the market simulator (during the analysis phase) and are forced to associate only a single level from each attribute with each product definition. Consider the following attribute with three levels:

One Three-Level Attribute
Sunroof
Extended Warranty
Global Positioning System (GPS)

This formulation does not permit simulating preference for a car that has both a sunroof and a GPS. Similarly, we could not simulate preference for an automobile that had none of these features because there is no *none* level. There are two ways to resolve this quandary. We can create an attribute with all potential combinations of these features. This results in an eight-level attribute, if you include the option that none of these features is available:

One Eight-Level Attribute
Sunroof, Extended Warranty, GPS
Sunroof, Extended Warranty, No GPS
Sunroof, No Extended Warranty, GPS
Sunroof, No Extended Warranty, No GPS
No Sunroof, Extended Warranty, GPS
No Sunroof, Extended Warranty, No GPS
No Sunroof, No Extended Warranty, GPS
No Optional Features

Or we can formulate three separate attributes each with two levels as follows:

Sunroof
No Sunroof
Sunroof

Extended Warranty
No Extended Warranty
Extended Warranty

GPS
No GPS
GPS

(Note that it is not necessary to state explicitly, for example, "No GPS." We could leave this level blank, so nothing would appear in that attribute position on the product profile.)

Creating an attribute with eight levels adds seven parameters to the model (see chapter 8 to understand why it is not eight) and forces the measurement of an explicit three-way interaction. With this more complex model definition, we can investigate whether there are diminishing returns (or unexpected synergies) by bundling the features. Splitting the options into three distinct binary attributes adds only three parameters if interaction parameters are not estimated.

Balance and Limit the Number of Attribute Levels

The number of levels you use to define an attribute can have a significant bearing on the results. The first concern has been called the number-of-levels effect (Currim, Weinberg, and Wittink 1981). All else being equal, attributes defined by more levels tend to get more importance. There is a large body of literature on this subject, and researchers recognize that both psychological and algorithmic effects play a role in the number-of-levels effect. The number-of levels effect is less problematic in ACA than full-profile conjoint methods such as CBC and traditional full-profile conjoint.

Ideally, you should try to balance the number of levels across attributes, especially for quantitative attributes such as price, speed, and weight. But there are situations in which some attributes in the real world (such as brand) have many more levels than other attributes. There is an argument for making the conjoint task mimic reality rather than balancing the number of levels and sacrificing realism. If in reality there are many more brands available on the shelf than package sizes, perhaps the same number-of-levels effect that influences conjoint analysis

results also influences real world choices. Buyers may pay more attention to brand variation than to variation in other attributes.

Another guideline is that you limit the number of levels on which quantitative attributes are described. For most purposes, you should not need more than about five levels to describe attributes such as price or speed. It's usually better to have more data at each price point than to have thinner measurements at more price points. Measuring too many points along a quantitative function can result in imprecise part-worths and troublesome reversals. If you cover the entire range of interest with fewer levels, you can interpolate between levels within the market simulator to get finer granularity if needed.

6.3 Use Prohibitions Sparingly

When we impose prohibitions, we ensure that certain levels of one attribute never appear with certain levels of another attribute. Prohibitions or prohibiting pairs should be used sparingly or not at all. Specifying unnecessary or excessive prohibitions is a common mistake in conjoint studies.

The problem usually begins when the analyst or client notices that some product combinations displayed during the interview are not realistic, given what currently exists in the market. Sometimes a product is shown with all the best features at the lowest price, or two attribute levels that would not naturally occur in the real world are paired together. The inclination is simply to prohibit such combinations.

It is not uncommon for experienced conjoint researchers to prohibit a select few levels from occurring with others. Such steps can lead to more realistic interviews. However, the wrong pattern of prohibitions or too many of prohibitions can be detrimental and sometimes fatal. Software for designing conjoint experiments includes efficiency diagnostics to assess whether prohibitions will pose problems for the precise estimation of respondent preferences (part-worth utilities). There are strategies for handling illogical or unnatural combinations of levels that can be less detrimental to design efficiency than prohibitions. These strategies include collapsing attributes and alternative-specific attributes.

There are other strategies for dealing with prohibitions. Consider an example with three brands of soda (e.g., Sawtooth Spritz, Kong Kola, and Martian Mist) and two package types (e.g., two-liter bottle or six-pack of twelve-ounce cans). Suppose that Martian Mist is only available in six-packs of twelve-ounce cans, and you are displaying actual pictures of the products, not potential products.

Rather than define a prohibition between Martian Mist and the two-liter bottle, it would make more sense to combine these two attributes into a single composite attribute with five levels, as illustrated in exhibit 6.1. Using the single-attribute approach, no prohibitions are required, but you will not be able to assess easily brand and package type as separate effects. This is probably not an issue if market simulations are used as the primary method of presenting results.

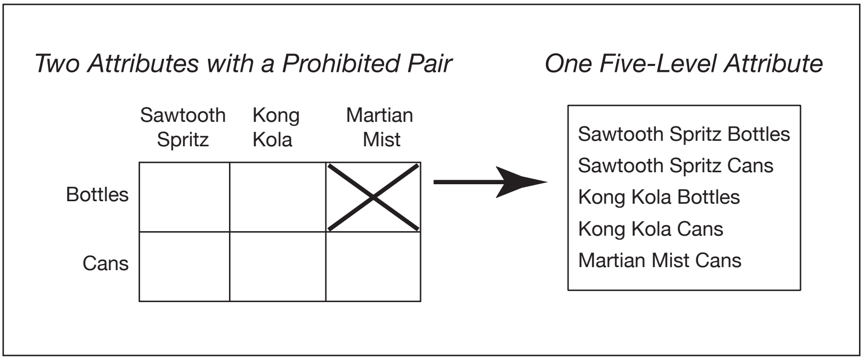


Exhibit 6.1. Resolving prohibitions with the composite-attribute approach

In the face of moderate to severe prohibitions, Adaptive Conjoint Analysis and other partial-profile conjoint methods are more robust than full-profile conjoint and full-profile choice-based conjoint methods. With full-profile methods, if two attributes have prohibited levels, the levels that appear for one attribute are necessarily correlated with the levels that appear for the other attribute. But with partial-profile methods, the two attributes do not always appear together within every conjoint question (one attribute may be missing). For such conjoint questions, those prohibitions have no effect, leading to relatively stable estimates of the levels for the two prohibited attributes.

6.4 Alternative-Specific Attributes

Alternative-specific attributes are a common and flexible way to deal with situations that would seem to require extreme prohibitions. Traditional conjoint analysis (CVA) and Adaptive Conjoint Analysis (ACA) use a common list of attributes to construct product profiles for the respondent. But CBC and ACBC approaches can employ attribute lists that are specific to the alternatives or concepts in the choice task.

Consider alternative ways of getting to work: car or bus. The attributes that are associated with car travel (cost of gas and parking) are different from the attributes associated with bus travel (wait times, cost per round trip, WiFi availability). Nonetheless, we might want respondents to choose between cars and buses within a CBC task. Rather than thinking about prohibitions between levels

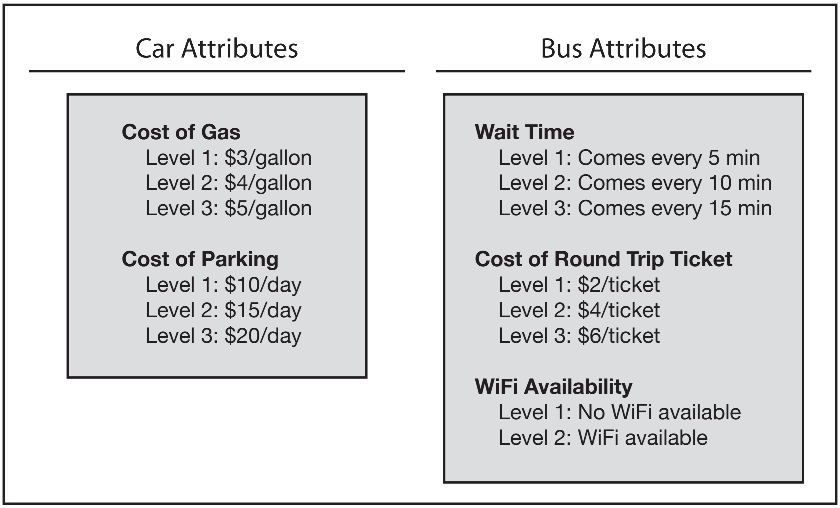


Exhibit 6.2. Alternative-specific attributes

of attributes to reflect the contingencies of traveling by car or bus, we can employ separate lists of attributes for the alternatives in the choice task.

As shown in exhibit 6.2, the car alternative is composed of car-travel attributes on the left-hand side, and the bus alternative is composed of bus-travel attributes on the right-hand side. Additional alternatives for trains, walking, or other modes of transportation could be added, each mode with is own specific attribute list.

Traditional goals of experimental design apply to studies with alternative-specific attributes. We may strive for design balance, so that each level of each attribute appears an equal number of times and each level appears an equal number of times with every other level of other attributes. But we do not have to worry about prohibiting specific levels of bus wait-times with levels of car gasoline costs because car and bus attributes are distinct. An example CBC task for a study with alternative-specific attributes is shown in exhibit 6.3.

Alternative-specific designs are not used with traditional conjoint (CVA) and Adaptive Conjoint Analysis (ACA). But, since the inception of choice-based conjoint studies in the 1970s, leading academics have employed alternative-specific designs. The flexibility of these designs is yet another reason why CBC is the most commonly used conjoint methodology today. Among Sawtooth Software users surveyed in 2012, 26 percent of CBC studies employed alternative-specific designs (Sawtooth Software 2012).

CBC designs can also include a mix of alternative-specific and generic attributes. For example, another attribute entitled “time to get to work” with levels 30 minutes, 40 minutes, and 60 minutes could be added to both the car and bus alternatives in exhibit 6.3.

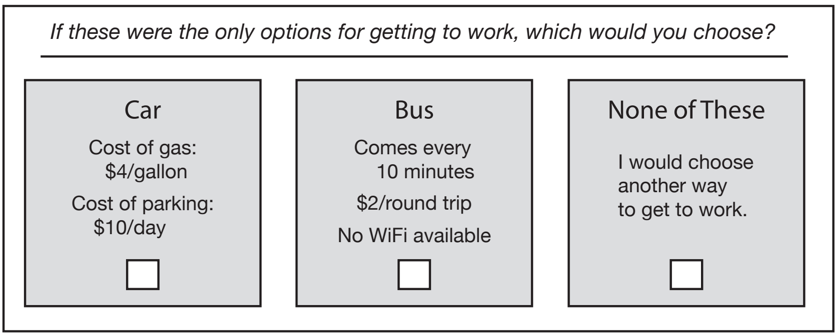


Exhibit 6.3. Alternative-specific choice task

Chapter 7

Sample Size Issues for Conjoint Analysis

"I'm about to conduct a conjoint analysis study. How large a sample size do I need? What will be the margin of error of my estimates if I use a sample of only 100 respondents?" These are common questions. Unfortunately, they are difficult questions to answer because many issues come into play:

- What is it exactly that you are trying to measure to get a statistically significant result: a specific part-worth, preference for a product, or the difference in preference between groups of people?
- Do you expect that the differences between features/products/groups you are trying to detect are subtle or strong?
- What level of certainty do you need to be able to act upon your conclusions: 99% confidence, 90% confidence, or what?
- How large is the total population in the market for your product?
- What conjoint methodology do you plan to use? How many conjoint questions will each respondent answer?
- Do you need to compare subsets of respondents, or are you going to be looking at results only as a whole?
- How homogenous is your market? Do people tend to think alike, or are there strong differences in preferences among individuals?
- How do you plan to select your sample? Will it be a random sample or convenience sample?
- How large of a budget do you have for the project?

Answers to these questions play a role in determining the appropriate sample size for a conjoint study. This chapter provides advice and tools to help conjoint researchers make sample size decisions. It involves more statistical theory and formulas than other chapters, so please bear with me.

Though most of the principles that influence sample size determination are based on statistics, successful researchers develop heuristics for quickly determining sample sizes based on experience, rules-of-thumb, and budget constraints. Let us begin our discussion by making a distinction between sampling and measurement error. Subsequent sections will discuss each of these sources of error.

7.1 Sampling Error versus Measurement Error

Errors are deviations from truth. In marketing research we are always concerned with reducing error in cost-effective ways. Assuming that you have selected the appropriate modeling method, there are two main sources of error that cause preference data to deviate from truth. The first is sampling error.

Sampling error occurs when samples of respondents deviate from the underlying population. If we have drawn a random sample (each population element has an equal probability of being selected), sampling error is due to chance. If, on the other hand, our sample is not random (for example, a convenience sample), the sampling errors may be systematic. With random sampling, we reduce sampling error by simply increasing the sample size. With nonrandom sampling, however, there is no guarantee that increasing sample size will make the samples more representative of the population.

To illustrate sampling error, assume we wanted to figure out how far the average adult can throw a baseball. If we drew a random sample of thirty people, and by chance happened to include Ichiro Suzuki (outfielder for the New York Yankees), our estimate would likely be farther than the true distance for the average adult. It is important to note that the samples we use in marketing research are rarely random. Some respondents resist being interviewed and, by selecting themselves out of our study, are a source of nonresponse bias.

A second source of error in conjoint data is measurement error. We reduce measurement error by having more or better data from each respondent. Consider again the example of the baseball toss. Suppose you are one of the study participants. You throw the ball, but you accidentally step into an uneven spot on the ground, and the ball does not go as far as you typically could throw it. If we asked you to take another few tosses, and averaged the results, we would reduce the measurement error and get a better idea of how far you could throw a baseball.

In conjoint analysis, we reduce measurement error by including more conjoint questions. We recognize, however, that respondents get tired, and there is a limit beyond which we can no longer get reliable responses, and therefore a limit to the amount we can reduce measurement error.

7.2 Binary Variables and Proportions

Sampling error is expressed in terms of standard errors, confidence intervals, and margins of error. We can begin to understand what these terms mean by considering binary variables and proportions. In fact, we will spend a good deal of time talking about confidence intervals for proportions because the statistical principles can be applied to choice-based conjoint results and shares of choice in market simulations for all conjoint techniques.

A binary variable is a categorical variable with exactly two levels, such as a yes/no item on a consumer survey or a true/false checklist item. Many product attributes in conjoint studies have exactly two levels. And consumer choice itself is binary—to choose or not, to buy or not. Binary variables are usually coded as 1 for yes and 0 for no. Looking across a set of binary variables, we see a set of 1s and 0s. We can count the number of 1s, and we can compute the proportion of 1s, which is the number of 1s divided by the sample size n.

In statistical theory, the sampling distribution of the proportion is obtained by taking repeated random samples from the population and computing the proportion for each sample. The standard error of the proportion is the standard deviation of these proportions across the repeated samples. The standard error of a proportion is given by the following formula:

$$\textbf{standard error of a proportion} = \sqrt{\frac{pq}{(n-1)}}$$

where p is the sample estimate of the proportion in the population, $q = (1 - p)$, and n is the sample size.

Most of us are familiar with the practice of reporting the results of opinion polls. Typically, a report may say something like this: "If the election were held today, Mike Jackson is projected to capture 50 percent of the vote. The survey was conducted by the XYZ company and has a margin of error of ± 3 percent." What is margin of error?

Margin of error refers to the upper and lower limits of a confidence interval. If we use what is known as the normal approximation to the binomial, we can obtain upper and lower limits of the 95% confidence interval for the proportion as

$$\textbf{margin of error for a proportion} = \pm 1.96\sqrt{\frac{pq}{(n-1)}}$$

Going back to the polling report from XYZ company, we note that margin of error has a technical meaning in classical statistics. If XYZ were to repeat the poll a large number of times (with a different random sample each time), 95 percent of the confidence intervals associated with these samples would contain the true proportion in the population. But, of course, 5 percent of the confidence intervals would not contain the true proportion in the population. Confidence intervals are random intervals. Their upper and lower limits vary from one sample to the next.

Suppose we interview 500 respondents (drawn using scientific random sampling) and ask whether they approve of the president's job performance, and suppose 65 percent say yes. What would be the margin of error of this statistic? We would compute the interval as follows:

$$\pm 1.96\sqrt{\frac{(0.65)(0.35)}{(500-1)}} = \pm 0.042$$

The margin of error is ± 4.2 percent for a confidence interval from 60.8 to 69.2 percent. We expect 95 percent of the confidence intervals constructed in this way to contain the true value of the population proportion.

Note that the standard error of the proportion varies with the size of the population proportion. So when there is agreement among people about a yes/no question on a survey, the value of p is closer to one or zero, and the standard error of the proportion is small. When there is disagreement, the value of p is closer to 0.50, and the standard error of the proportion is large. For any given sample size n, the largest value for the standard error occurs when $p = 0.50$.

When computing confidence intervals for proportions, then, the most conservative approach is to assume that the value of the population proportion is 0.50. That is, for any given sample size and confidence interval type, $p = 0.50$ will provide the largest standard error and the widest margin of error. Binary variables and proportions have this special property—for any given sample size n and confidence interval type, we know the maximum margin of error before we collect the data. The same cannot be said for continuous variables, which we discuss in the next section.

7.3 Continuous Variables and Means

With continuous variables (ratings-based responses to conjoint profiles), one cannot estimate the standard error before fielding a study. The standard error of the mean is directly related to the standard deviation of the continuous variable, which differs from one study to the next and from one survey question to the next. Assuming a normal distribution, the standard error of the mean is given by

$$\textbf{standard error of the mean} = \frac{\textbf{standard deviation}}{\sqrt{n}}$$

And the margin of error associated with a 95% confidence interval for the mean is given by

$$\textbf{margin of error for the mean} = \pm 1.96\big(\textbf{standard error of the mean}\big)$$

Suppose we had conducted an ACA study with forty respondent interviews. We want to estimate purchase likelihood for a client's planned product introduction with a margin of error of ± 3 and a 95% confidence level. We run an ACA market simulation to estimate purchase likelihood on a 100-point scale, and the simulator reports the standard error next to the purchase likelihood estimate:

Total Respondents = 40

	Purchase Likelihood	*Standard Error*
Product A	78.34	3.06

The margin of error is $\pm 1.96 \times 3.06 = \pm 6.00$, so we need to cut the margin of error in half to achieve our ± 3 target level of precision. We know that the standard error of the mean is equal to the standard deviation divided by the square-root of the sample size. To decrease the standard error by a factor of two, we must increase sample size by a factor of four. Therefore, we need to interview about $40 \times 4 = 160$ or 120 additional respondents to obtain a margin of error of ± 3 for purchase likelihood.

7.4 Small Populations and the Finite Population Correction

The examples we have presented thus far have assumed infinite or very large populations. But suppose that, instead of estimating the job performance rating of the president by the United States population at large, we wanted to estimate (with a margin of error of ± 3 percent) the job performance rating of a school principal by members of the PTA. Suppose there are only 100 members of the PTA. How many PTA members do we need to interview to achieve a margin of error of ± 3 percent for our estimate?

First, we introduce a new term: finite population correction. The formula for the finite population correction is $\frac{(N-n)}{(N-1)}$, where n is the sample size and N is the population size. The formula for the finite population correction is often simplified to $(1-f)$, where $f = \frac{n}{N}$, which is approximately equivalent to $\frac{(N-n)}{(N-1)}$ for all except the smallest of populations.

After a population reaches about 5,000 individuals, one can generally ignore the finite population correction factor because it has a very small impact on sample size decisions. Using the simplified finite population correction for a finite sample, the margin of error for a proportion and a 95% confidence interval is equal to

$$\pm 1.96 \sqrt{(1-f)\frac{pq}{(n-1)}}$$

The finite population correction may also be used for continuous variables and means.

With a population of 100, we can solve for n assuming an expected proportion. The worst-case scenario (i.e., the one that has the largest standard error) is for a 0.50 proportion, so it is standard to let $p = 0.50$. Solving for n, we discover that we would need to interview 92 PTA members, or 92 percent of the population to achieve a margin of error of ± 3 percent.

The important point to be made is that with small populations, you may have to interview a significant proportion of the population to achieve stable estimates. Suppose your client produces a very expensive, highly specialized piece of machinery, for which there were only 100 total potential customers in the world. Given many people's unwillingness to complete surveys, it will likely be much more difficult to complete surveys with 92 out of 100 potential buyers of this product than to interview, say, 1,000 potential buyers of something like office chairs, for which there are so many buyers as to approximate an infinite population. Even so, in terms of estimating a proportion, both scenarios lead to the same margin of error when projecting to the population of interest.

Conjoint studies may be used for large or small populations. We can use conjoint analysis for even the smallest of populations, provided we interview enough respondents to represent the population adequately.

7.5 Measurement Error in Conjoint Studies

Many researchers and dozens of data sets have demonstrated that conjoint utilities do a good job of predicting individual respondents' preferences for products. Holdout choice sets (choice tasks not used to estimate utilities) are often included in conjoint questionnaires. Using the conjoint data, a respondent's holdout choices usually can be predicted with a hit rate of roughly 75 to 85 percent. These choice tasks typically include between three and five different product concepts, so by chance we expect a success rate between 20 and 33 percent.

The hit rates with conjoint are significantly greater than chance and significantly better than the marketer's best guesses—even if the marketer knows each customer very well. In fact, conjoint predictions at the individual level frequently approach or sometimes even exceed test-retest reliability, suggesting that a good set of conjoint utilities is about as reliable at predicting choices to repeated holdout tasks as the respondents' earlier choices.

If there were only one buyer of your product in the world, you could learn a great deal about that individual's preferences from a conjoint interview. The utility data would be reasonably accurate for predicting his or her preferences and weights placed upon attributes. We can learn a great deal about an individual respondent provided we ask that respondent the right questions and enough questions. Let us consider numbers of conjoint questions or tasks needed for alternative methods of conjoint analysis.

Adaptive Conjoint Analysis

An Adaptive Conjoint Analysis (ACA) interview results in a set of utilities for each individual. We want conjoint measurements for each individual in the study to be as accurate as possible.

Of the conjoint methods discussed in this book, ACA is perhaps the best at reducing measurement error. ACA's interviews adapt to the respondent, asking questions designed to be maximally relevant and efficient for refining utility estimates. The priors section helps in stabilizing the utility estimates at the individual level. One sees fewer reversals in part-worths (out-of-order utilities) for ordered attributes like price in ACA than in traditional conjoint and choice-based conjoint with individual estimation.

In ACA, one needs to decide how many pairs questions to ask. The number of pairs each respondent completes plays a significant role in reducing measurement error. The suggested number of pairs is $3(K - k - 1) - K$, where K is the total number of levels across all attributes and k is number of attributes. If respondents answer as many pairs as suggested, a total of three times the number of observations as parameters are available at the individual level for computing utilities (this includes information from the self-explicated priors). Sometimes the suggested number of pairs is greater than respondents can reasonably do. You should make sure not to overburden respondents because this can lead to poor results. You can ask fewer than the recommended number of pairs, though this increases the measurement error for each individual.

If your sample size is particularly small and the number of attributes to measure is large, ACA or ACBC may be better tools to use. In fact, it is possible to have an entire research study designed to learn about the preferences of one respondent, such as an important buyer of an expensive industrial product. As we discussed in chapter 5, there are many considerations for determining whether ACA is appropriate for a study. For further discussion of ACA measurement, estimation, and sample size issues, see Johnson (1987a).

Traditional Conjoint Studies

Like ACA, traditional full-profile conjoint (such as Sawtooth Software's CVA or SPSS's conjoint module) usually leads to the estimation of individual-level part-worth utilities. Again, the minimum sample size is one. But, because the traditional conjoint methodology does not include a self-explicated priors section, its utilities tend to have greater variability (larger standard errors) at the individual level relative to ACA (holding respondent effort equal).

One should include enough conjoint questions or cards to reduce measurement error sufficiently. Sawtooth Software's CVA manual suggests asking enough questions to obtain three times the number of observations as parameters to be estimated, or a number equal to $3(K - k + 1)$, where K is the total number of levels across all attributes and k is the number of attributes.

Respondents sometimes lack the energy or patience to answer many questions. We need to strike a good balance between overworking the respondent (and getting noisy data) and not asking enough questions to stabilize the estimates.

Choice-Based Conjoint

Though generally considered more realistic than traditional conjoint, choice-based questions are a relatively inefficient way to learn about preferences. As a result, sample sizes are typically larger than with ACA or traditional ratings-based conjoint, and choice-based conjoint (CBC) results have traditionally been analyzed by aggregating respondents. Since the late 1990s, hierarchical Bayes has permitted individual-level estimation of part-worth utilities from CBC data. But to compute individual-level models, HB uses information from many respondents to refine the utility estimates for each individual. Therefore, one usually does not calculate utilities using a sample size of one. It should be noted, however, that logit analysis can be run at the individual level, if the number of parameters to be estimated is small, the design is highly efficient, and the number of tasks is large.

There are rules-of-thumb for determining sample sizes for CBC if we are willing to assume aggregate estimation of effects. Like proportions, choices reflect binary data, and the rules for computing confidence intervals for proportions are well defined and known prior to collecting data.

Consider a design with three brands and three prices. Assume each person completes ten tasks, and each task displays three products (i.e., each brand and price occurs once per task). If we interview 100 respondents, each brand will have been available for choice

$$(100\text{ respondents}) \times (10\text{ tasks}) \times \frac{(3\text{ concepts})}{(3\text{ brands})} = 1,000\text{ times}$$

Johnson and Orme (1996) looked at about twenty commercial choice-based conjoint data sets and determined that having each respondent complete ten tasks is about as good at reducing error as having ten times as many respondents complete one task. Of course, in the limit this suggestion is ridiculous. It does not make sense to say that having one respondent complete 1,000 tasks is as good as having 1,000 respondents complete one task. But, according to Johnson and Orme (1996) simulation results, if a researcher obtains data from three to four hundred respondents, doubling the number of tasks they complete is about as good (in terms of reducing overall error) as doubling the sample size. It makes sense from a cost-benefit standpoint, then, to have respondents complete many choice tasks.

Johnson, who is the author of Sawtooth Software's CBC System, has recommended a rule-of-thumb when determining minimum sample sizes for aggregate-level full-profile CBC modeling: set

$$\frac{nta}{c} \geq 500$$

where n is the number of respondents, t is the number of tasks, a is number of alternatives per task (not including the *none* alternative), and c is the number of analysis cells. When considering main effects, c is equal to the largest number of levels for any one attribute. If you are also considering all two-way interactions, c is equal to the largest product of levels of any two attributes (Johnson and Orme 2003).

Over the years, we have become concerned that practitioners use Johnson's rule-of-thumb to justify sample sizes that are too small. Some feel that they will have ample stability in estimates when each main-effect level of interest is represented across the design about 500 times. But 500 was intended to be a minimum threshold when researchers cannot afford to do better. It would be better, when possible, to have 1,000 or more representations per main-effect level.

7.6 Typical Sample Sizes and Practical Guidelines

The recommendations below assume infinite or very large populations. They are based on the theories above and our observations of common practices in the market research community:

- Sample sizes for conjoint studies generally range from about 150 to 1,200 respondents.
- If the purpose of your research is to compare groups of respondents and detect significant differences, you should use a large enough sample size to accommodate a minimum of about 200 per group. Therefore, if you are conducting a segmentation study and plan to divide respondents into as many as four groups (i.e., through cluster analysis) it would be wise to include, at a minimum, $4 \times 200 = 800$ respondents. This, of course, assumes your final group sizes will be about equal, so one would usually want more data. The stronger segmentation studies include about 800 or more respondents.
- For robust quantitative research where one does not intend to compare subgroups, I would recommend at least 300 respondents. For investigational work and developing hypotheses about a market, between thirty and sixty respondents may do.

These suggestions have to be weighed against research costs. There are difficult decisions to be made based on experience, the application of statistical principles, and sound judgment. If, after the fact, you find yourself questioning whether you really needed to have collected such a large sample size for a particular project, it is an interesting exercise to delete a random subset of the data to see how having fewer respondents would have affected your findings.

A thorough discussion of sampling and measurement errors would require more time and many more pages. The reader is encouraged to consult other sources in these areas. For statistics and sampling see Snedecor and Cochran (1989) and Levy and Lemeshow (1999). For measurement theory see Nunnally (1967).

Chapter 8

Traditional Conjoint Analysis with Excel

A traditional conjoint analysis may be thought of as a multiple regression problem. The respondent's ratings for the product concepts are observations on the dependent variable. The characteristics of the product or attribute levels are observations on the independent or predictor variables. The estimated regression coefficients associated with the independent variables are the part-worth utilities or preference scores for the levels. The R^2 for the regression characterizes the internal consistency of the respondent.

Consider a conjoint analysis problem with three attributes, each with levels as follows:

Brand	*Color*	*Price*
A	Red	$50
B	Blue	$100
C		$150

For simplicity, let us consider a full-factorial experimental design. A full-factorial design includes all possible combinations of the attributes. There are 18 possible product concepts or cards that can be created from these three attributes:

$$3 \text{ brands} \times 2 \text{ colors} \times 3 \text{ prices} = 18 \text{ cards}$$

Further assume that respondents rate each of the 18 product concepts on a scale from 0 to 10, where 10 represents the highest degree of preference. Exhibit 8.1 shows the experimental design.

We can use Microsoft Excel to analyze data from traditional conjoint questionnaires. This chapter shows how to code, organize, and analyze data from one hypothetical respondent, working with spreadsheets and spreadsheet functions. Multiple regression functions come from the Excel Analysis ToolPak add-in.

Card	Brand	Color	Price ($)
1	A	Red	50
2	A	Red	100
3	A	Red	150
4	A	Blue	50
5	A	Blue	100
6	A	Blue	150
7	B	Red	50
8	B	Red	100
9	B	Red	150
10	B	Blue	50
11	B	Blue	100
12	B	Blue	150
13	C	Red	50
14	C	Red	100
15	C	Red	150
16	C	Blue	50
17	C	Blue	100
18	C	Blue	150

Exhibit 8.1. Full-factorial experimental design

8.1 Data Organization and Coding

Assume the data for one respondent have been entered into an Excel spreadsheet, illustrated in exhibit 8.2. The first card is made up of the first level on each of the attributes: (Brand A, Red, $50). The respondent rated that card a 5 on the preference scale. The second card has the first level on brand and color and the second level on price: (Brand A, Red, $100). This card gets a 5 on the preference scale. And so on.

After collecting the respondent data, the next step is to code the data in an appropriate manner for estimating utilities using multiple regression. We use a procedure called dummy coding for the independent variables or product characteristics. In its simplest form, dummy coding uses a 1 to reflect the presence of a feature, and a 0 to represent its absence. The brand attribute would be coded as three separate columns, color as two columns, and price as three columns. Applying dummy coding results in an array of columns as illustrated in exhibit 8.3. Again, we see that card 1 is defined as (Brand A, Red, $50), but we have expanded the layout to reflect dummy coding.

To this point, the coding has been straightforward. But there is one complication that must be resolved. In multiple regression analysis, no independent variable may be perfectly predictable based on the state of any other independent variable or combination of independent variables. If so, the regression procedure cannot separate the effects of the confounded variables. We have that problem with the data as coded in exhibit 8.3, since, for example, we can perfectly predict the state of Brand A based on the states of Brand B and Brand C. This situation is called linear dependency.

To resolve this linear dependency, we omit one column from each attribute. It really doesn't matter which column (level) we drop, and for this example we have excluded the first level for each attribute, to produce a modified data table, as illustrated by exhibit 8.4.

Even though it appears that one level from each attribute is missing from the data, they are really implicitly included as reference levels for each attribute. The explicitly coded levels are estimated as contrasts with respect to the omitted levels, which are constrained to have a weight of 0.

	A	B	C	D	E
1	Card	Brand	Color	Price	Preference
2	1	1	1	50	5
3	2	1	1	100	5
4	3	1	1	150	0
5	4	1	2	50	8
6	5	1	2	100	5
7	6	1	2	150	2
8	7	2	1	50	7
9	8	2	1	100	5
10	9	2	1	150	3
11	10	2	2	50	9
12	11	2	2	100	6
13	12	2	2	150	5
14	13	3	1	50	10
15	14	3	1	100	7
16	15	3	1	150	5
17	16	3	2	50	9
18	17	3	2	100	7
19	18	3	2	150	6

Exhibit 8.2. Excel spreadsheet with conjoint data

	H	I	J	K	L	M	N	O	P	Q
1	Card	A	B	C	Red	Blue	$50	$100	$150	Preference
2	1	1	0	0	1	0	1	0	0	5
3	2	1	0	0	1	0	0	1	0	5
4	3	1	0	0	1	0	0	0	1	0
5	4	1	0	0	0	1	1	0	0	8
6	5	1	0	0	0	1	0	1	0	5
7	6	1	0	0	0	1	0	0	1	2
8	7	0	1	0	1	0	1	0	0	7
9	8	0	1	0	1	0	0	1	0	5
10	9	0	1	0	1	0	0	0	1	3
11	10	0	1	0	0	1	1	0	0	9
12	11	0	1	0	0	1	0	1	0	6
13	12	0	1	0	0	1	0	0	1	5
14	13	0	0	1	1	0	1	0	0	10
15	14	0	0	1	1	0	0	1	0	7
16	15	0	0	1	1	0	0	0	1	5
17	16	0	0	1	0	1	1	0	0	9
18	17	0	0	1	0	1	0	1	0	7
19	18	0	0	1	0	1	0	0	1	6

Exhibit 8.3. Excel spreadsheet with coded data

	S	T	U	V	W	X	Y
1	Card	B	C	Blue	$100	$150	Preference
2	1	0	0	0	0	0	5
3	2	0	0	0	1	0	5
4	3	0	0	0	0	1	0
5	4	0	0	1	0	0	8
6	5	0	0	1	1	0	5
7	6	0	0	1	0	1	2
8	7	1	0	0	0	0	7
9	8	1	0	0	1	0	5
10	9	1	0	0	0	1	3
11	10	1	0	1	0	0	9
12	11	1	0	1	1	0	6
13	12	1	0	1	0	1	5
14	13	0	1	0	0	0	10
15	14	0	1	0	1	0	7
16	15	0	1	0	0	1	5
17	16	0	1	1	0	0	9
18	17	0	1	1	1	0	7
19	18	0	1	1	0	1	6

Exhibit 8.4. Modified data table for analysis with Excel

8.2 Multiple Regression Analysis

Microsoft Excel offers a simple multiple regression tool (under Tools + Data Analysis + Regression with the Analysis Toolpak add-in installed). Using the tool, we can specify the preference score (column Y) as the dependent variable (Input Y Range) and the five dummy-coded attribute columns (columns T through X) as independent variables (Input X range). You should also make sure a constant is estimated; this usually happens by default (by not checking the box labeled “Constant is zero”).

The mathematical expression of the model is as follows:

$$Y = b_0 + b_1(\text{Brand B}) + b_2(\text{Brand C}) + b_3(\text{Blue}) + b_4(\$100) + b_5(\$150) + e$$

where Y is the respondent’s preference for the product concept, b_0 is the constant or intercept term, b_1 through b_5 are beta weights (part-worth utilities) for the features, and e is an error term. In this formulation of the model, coefficients for the reference levels are equal to 0. The solution minimizes the sum of squares of the errors in prediction over all observations.

A portion of the output from Excel is illustrated in exhibit 8.5. Using that output (after rounding to two decimal places of precision), the utilities (coefficients) are the following:

Brand	*Color*	*Price*
A = 0.00	Red = 0.00	\$ 50 = 0.00
B = 1.67	Blue = 1.11	\$100 = -2.17
C = 3.17		\$150 = -4.50

The constant or intercept term is 5.83, and the fit for this respondent $R^2 = 0.90$. Depending on the consistency of the respondent, the fit values range from a low of 0 to a high of 1.0. The standard errors of the regression coefficients (betas) reflect how precisely we are able to estimate those coefficients with this design. Lower standard errors are better. The remaining statistics presented in Excel’s output are beyond the scope of this chapter and are generally not of much use when considering individual-level conjoint analysis problems.

Most traditional conjoint analysis problems solve a separate regression equation for each respondent. Therefore, to estimate utilities, the respondent must have evaluated at least as many cards as parameters to be estimated. When the respondent answers the minimum number of conjoint cards to enable estimation, this is called a saturated design. While such a design is easiest on the respondent, it leaves no room for respondent error. It also always yields an R^2 of 1, and therefore no ability to assess respondent consistency.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.94890196
R Square	0.90041494
Adjusted R Sq	0.85892116
Standard Erro	0.94280904
Observations	18

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	96.4444444	19.2888889	21.7	1.2511E-05
Residual	12	10.6666667	0.88888889		
Total	17	107.111111			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5.83333333	0.54433105	10.7165176	1.6872E-07
X Variable 1	1.66666667	0.54433105	3.06186218	0.00986485
X Variable 2	3.16666667	0.54433105	5.81753814	8.2445E-05
X Variable 3	1.11111111	0.44444444	2.5	0.0279154
X Variable 4	-2.16666667	0.54433105	-3.98042083	0.0018249
X Variable 5	-4.5	0.54433105	-8.26702788	2.6823E-06

Exhibit 8.5. Conjoint analysis with multiple regression in Excel

One can easily determine the number of parameters to be estimated in a traditional conjoint analysis:

$$\#\,\textbf{parameters to be estimated} = (\#\,\textbf{levels}) - (\#\,\textbf{attributes}) + 1$$

Most good conjoint designs in practice include more observations than parameters to be estimated (usually 1.5 to 3 times more). The design we have featured in this chapter has three times as many cards (observations) as parameters to be estimated. These designs usually lead to more stable estimates of respondent utilities than saturated designs.

Only in the smallest of problems (such as our example with three attributes and eight total levels) would we ask people to respond to all possible combinations of attribute levels. Large full-factorial designs are not practical. Fortunately, design catalogs and computer programs are available to find efficient fractional-factorial designs. Fractional-factorial designs show an efficient subset of the possible combinations and provide enough information to estimate utilities.

In our worked example, the standard errors for the color attribute are lower than for brand and price (recall that lower standard errors imply greater precision of the beta estimate). Because color only has two levels (as compared to three each for brand and price), each color level has more representation within the design. Therefore, more information is provided for each color level than is provided for the three-level attributes.

Chapter 9

Interpreting the Results of Conjoint Analysis

Conjoint analysis provides various outputs for analysis, including part-worth utilities, counts, importances, shares of preference, and purchase likelihood simulations. This chapter discusses these measures and gives guidelines for interpreting results and presenting findings to management.

Before focusing on conjoint data, it is useful to review some fundamentals for interpreting quantitative data. The discussion of the nature of measurement scales follows the classic discussion of Stevens (1946), which has been adopted by numerous social scientists and business researchers. For current definitions and discussion, one can refer to a book on business statistics (Lapin 1993).

9.1 Nature of Quantitative Data

There are four general types of quantitative data:

- Nominal data. Here the numbers represent categories, such as (1=male, 2=female) or (20=Italy, 21=Canada, 22=Mexico). It is not appropriate to perform mathematical operations such as addition or subtraction with nominal data or to interpret the relative size of the numbers.
- Ordinal data. These commonly occur in market research in the form of rankings. If a respondent ranks five brands from best 1 to worst 5, we know that a 1 is preferred to a 2. An example of an ordinal scale is the classification of strengths of hurricanes. A category 3 hurricane is stronger and more damaging than a category 2 hurricane. It is generally not appropriate to apply arithmetic operations to ordinal data. The difference in strength between a category 1 and a category 2 hurricane is not necessarily equal to the difference in strength between a category 2 and a category 3. Nor can we say that a category 2 is twice as strong as a category 1 hurricane.

I would like to express special thanks to Rich Johnson for his contributions to this chapter in the section entitled "Price Elasticity, Price Sensitivity, and Willingness to Pay."

- Interval data. These permit the simple operations of addition and subtraction. The rating scales so common to market research provide interval data. The Celsius scale is an example of an interval scale. Each degree of temperature represents an equal heat increment. It takes the same amount of heat to raise the temperature of a cup of water from 10 to 20 degrees as from 20 to 30 degrees. The zero point is arbitrarily tied to the freezing point of distilled water. Sixty degrees is not twice as hot as 30 degrees, and the ratio 60/30 has no meaning.
- Ratio data. These data permit all basic arithmetic operations, including division and multiplication. Examples of ratio data include weight, height, time increments, revenue, and profit. The zero point is meaningful in ratio scales. The difference between 20 and 30 kilograms is the same as the difference between 30 and 40 kilograms, and 40 kilograms is twice as heavy as 20 kilograms.

9.2 Conjoint Utilities

Conjoint utilities or part-worths are scaled to an arbitrary additive constant within each attribute and are interval data. The arbitrary origin of the scaling within each attribute results from dummy coding in the design matrix. We could add a constant to the part-worths for all levels of an attribute or to all attribute levels in the study, and it would not change our interpretation of the findings.

When using a specific kind of dummy coding called effects coding, utilities are scaled to sum to zero within each attribute. A plausible set of part-worth utilities for fuel efficiency measured in miles per gallon might look like this:

Fuel Efficiency	*Utility*
30 mpg	-1.0
40 mpg	0.0
50 mpg	1.0

30 mpg received a negative utility value, but this does not mean that 30 mpg was unattractive. In fact, 30 mpg may have been acceptable to all respondents. But, all else being equal, 40 mpg and 50 mpg are better. The utilities are scaled to sum to zero within each attribute, so 30 mpg must receive a negative utility value. Other kinds of dummy coding arbitrarily set the part-worth of one level within each attribute to zero and estimate the remaining levels as contrasts with respect to zero.

Whether we multiply all the part-worth utilities by a positive constant or add a constant to each level within a study, the interpretation is the same. Suppose we have two attributes with the following utilities:

Color	*Utility*	*Brand*	*Utility*
Blue	30	A	20
Red	20	B	40
Green	10	C	10

The increase in preference from Green to Blue (twenty points) is equal to the increase in preference between brand A and brand B (also twenty points). However, due to the arbitrary origin within each attribute, we cannot directly compare values between attributes to say that Red (twenty utiles) is preferred equally to brand A (twenty utiles). And even though we are comparing utilities within the same attribute, we cannot say that Blue is three times as preferred as Green (30/10). Interval data do not support ratio operations.

9.3 Counts

When using choice-based conjoint (CBC), the researcher can analyze the data by counting the number of times an attribute level was chosen relative to the number of times it was available for choice. In the absence of prohibitions, counts proportions are closely related to conjoint utilities. If prohibitions were used, counts are biased. Counts are ratio data. Consider the following counts proportions:

Color	*Proportion*	*Brand*	*Proportion*
Blue	0.50	A	0.40
Red	0.30	B	0.50
Green	0.20	C	0.10

We can say that brand A was chosen four times as often as brand C (0.40/0.10). But, as with conjoint utilities, we cannot report that Brand A is preferred to Red.

9.4 Attribute Importance

Sometimes we want to characterize the relative importance of each attribute. We can do this by considering how much difference each attribute could make in the total utility of a product. That difference is the range in the attribute's utility values. We calculate percentages from relative ranges, obtaining a set of attribute importance values that add to 100 percent, as illustrated in exhibit 9.1. For this respondent, whose data are shown in the exhibit, the importance of brand is 26.7 percent, the importance of price is 60 percent, and the importance of color is 13.3 percent. Importances depend on the particular attribute levels chosen for the study. For example, with a narrower range of prices, price would have been less important.

Attribute	*Level*	*Part-Worth Utility*	*Attribute Utility Range*	*Attribute Importance*
Brand	A B C	30 60 20	60 - 20 = 40	(40/150) x 100% = 26.7%
Price	$50 $75 $100	90 50 0	90 - 0 = 90	(90/150) x 100% = 60.0%
Color	Red Pink	20 0	20 - 0 = 20	(20/150) x 100% = 13.3%

Utility Range Total
40 + 90 + 20 = 150

Exhibit 9.1. Relative importance of attributes

When summarizing attribute importances for groups, it is best to compute importances for respondents individually and then average them, rather than computing importances from average utilities. For example, suppose we were studying two brands, Coke and Pepsi. If half of the respondents preferred each brand, the average utilities for Coke and Pepsi would be tied, and the importance of brand would appear to be zero.

Importance measures are ratio-scaled, but they are also relative, study-specific measures. An attribute with an importance of twenty percent is twice as important as an attribute with an importance of ten, given the set of attributes and levels used in the study. That is to say, importance has a meaningful zero point, as do all percentages. But when we compute an attribute's importance, it is always relative to the other attributes being used in the study. And we can compare one attribute to another in terms of importance within a conjoint study but not across studies featuring different attribute lists.

When calculating importances from CBC data, it is advisable to use part-worth utilities resulting from latent class (with multiple segments) or, better yet, HB estimation, especially if there are attributes on which respondents disagree about preference order of the levels. (Recall the previous Coke versus Pepsi example.)

One of the problems with standard importance analysis is that it considers the extremes within an attribute, irrespective of whether the part-worth utilities follow rational preference order. The importance calculations capitalize on random error, and attributes with very little to no importance can be biased upward in importance. There will almost always be a difference between the part-worth utilities of the levels, even if it is due to random noise alone. For that reason, many analysts prefer to use sensitivity analysis in a market simulator to estimate the impact of attributes.

9.5 Sensitivity Analysis Using Market Simulations

Conjoint part-worths and importances may be difficult for nonresearchers to understand. Many presentations to management go awry when the focus of the conversation turns to explaining how part-worths are estimated or, given the scaling resulting from dummy variable coding, how one can or cannot interpret part-worths.

We suggest using market simulators to make the most of conjoint data and to communicate the results of conjoint analysis. When two or more products are specified in the market simulator, we can estimate the percentage of respondents who would prefer each. The results of market simulators are easy to interpret because they are scaled from zero to one hundred. And, unlike part-worth utilities, simulation results (shares of preference) are assumed to have ratio scale properties—it is legitimate to claim that a 40 percent share of preference is twice as much as a 20 percent share. Sensitivity analysis using market simulation offers a way to report preference scores for each level of each product attribute.

The sensitivity analysis approach can show us how much we can improve (or make worse) a product's overall preference by changing its attribute levels one at a time, while holding all other attributes constant at base case levels. We usually conduct sensitivity analyses for products assuming no reaction by the competition. In this way, the impact of each attribute level is estimated within the specific and appropriate context of the competitive landscape. For example, the value of offering a round versus a square widget depends on both the inherent desirability (utility) of round and square shapes and how many current competitors are offering round or square shapes. (Note that if no relevant competition exists or if levels needed to describe competitors are not included in the study, then it is possible to conduct sensitivity simulations considering the strength of a single product concept versus the option of purchasing nothing, or considering the product's strength in terms of purchase likelihood.)

Conducting sensitivity analysis starts by simulating shares of choice among products in a base case market. Then, we change product characteristics one level at a time (holding all other attributes constant at base case levels), We run the market simulation repeatedly to capture the incremental effect of each attribute level upon product choice. After we test all levels within a given attribute, we return that attribute to its base case level prior to testing another attribute.

To illustrate the method, we consider an example involving a study of mid-range televisions in 1997. The attributes in the study were as follows:

Brand
Sony
RCA
JVC

Screen Size
25-inch
26-inch
27-inch

Sound Capability
Mono Sound
Stereo Sound
Surround Sound

Channel Block Capability
None
Channel Blockout

Picture-in-Picture Capability
None
Picture-in-Picture

Price
$300
$350
$400
$450

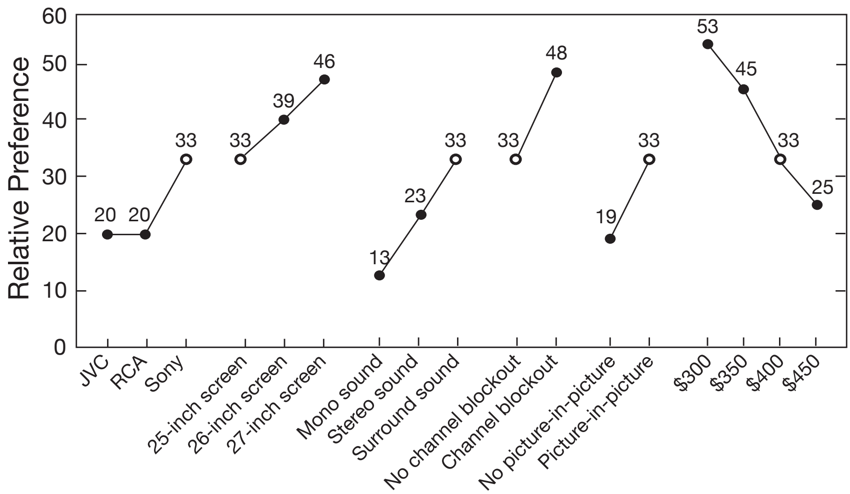


Figure 9.1. Results of sensitivity analysis

Suppose we worked for Sony and the competitive landscape was represented by this base case scenario (3 products, each represented by rows of features in the table below):

Brand	*Screen Size*	*Sound Capability*	*Channel Blockout Capability*	*Picture-in-Picture Capability*	*Price*
Sony	25-inch	Surround	None	Picture-in-Picture	$400
RCA	27-inch	Stereo	None	Picture-in-Picture	$350
JVC	25-inch	Stereo	None	None	$300

Let us assume that, for this base case scenario, the Sony product captured 33 percent relative share of preference.

For a market simulation, we can modify the Sony product to have other levels of screen size, sound capability, channel blockout capability, and picture-in-picture capability, while holding the products from RCA and JVC constant. Figure 9.1 shows estimated shares of preference from this type of market simulation or sensitivity analysis. The potential (adjacent-level) improvements to Sony's product can be ranked as follows:

- Add channel blockout (48 relative preference)
- Reduce price to $350 (45 relative preference)
- Increase screen size to 26-inch (39 relative preference)

Sony cannot change its brand to RCA or JVC, so the brand attribute is irrelevant to management decision making (except to note that the Sony brand is preferred to RCA and JVC). And, although it is unlikely that Sony would want to reduce its features and capabilities, we can observe a loss in relative preference by including levels of inferior preference. One of those is price. Increasing the price to $450 results in a lower relative preference of 25 percent.

Before making recommendations to Sony management, we would, of course, conduct more sophisticated what-if analyses, varying more than one attribute at a time. Nonetheless, the one-attribute-at-a-time approach to sensitivity analysis provides a good way to assess relative preferences of product attributes.

9.6 Price Elasticity, Price Sensitivity, and Willingness to Pay

The results of conjoint analysis may be used to assess the price elasticity of products and services. It may also be used to assess buyer price sensitivity and willingness to pay. To begin this section, we should define price elasticity, price sensitivity, and willingness to pay:

- Price elasticity, by which we mean the price elasticity of demand, is the percentage change in quantity demanded divided by the percentage change in price. Price elasticity relates to the aggregate demand for a product and the shape of the demand curve. It is a characteristic of a product in a market.
- Price sensitivity is a characteristic of buyers or consumers. Some people are more sensitive to price changes than others, and the degree to which they are price sensitive can vary from one product or service to the next, one market to the next, or one time to the next. It can also vary with the characteristics of products described in terms of product attributes.
- Willingness to pay is a characteristic of buyers or consumers. A measure of willingness to pay shows how much value an individual consumer places on a good or service. It is measured in terms of money.

Conjoint analysis is often used to assess how buyers trade off product features with price. Researchers can test the price sensitivity of consumers to potential product configurations using simulation models based on conjoint results. Most often a simulation is done within a specific context of competitors. But when a product is new to the market and has no direct competitors, price sensitivity of consumers for that new product can be estimated compared to other options such as buying nothing.

The common forms of conjoint analysis measure contrasts between levels within attributes. The part-worths of levels are estimated on an interval scale with an arbitrary origin, so the absolute magnitudes of utilities for levels taken alone have no meaning. Each attribute's utilities are determined only to within an arbitrary additive constant, so a utility level from one attribute cannot be directly compared to another from a different attribute. To a trained conjoint analyst, an

array of utilities conveys a clear meaning. But that meaning is often difficult for others to grasp. It is not surprising, then, that researchers look for ways to make conjoint utilities easier to interpret.

Monetary Scaling Trap

One common attempt to make conjoint utilities more understandable is to express them in monetary terms, or dollar equivalents. This is a way of removing the arbitrariness in their scaling. To do this, price must be included as an attribute in the conjoint design. Note that we cannot attach a monetary value to a single level (such as the color green), but must express the value in terms of differences between two colors, such as "green is worth $5 more than yellow." But if the attribute is binary (present/absent) such as "has sunroof" versus "doesn't have sunroof," the expressed difference is indeed the value of having the feature versus not having it.

The idea of converting utilities to dollar values can be appealing to managers. But some approaches to converting utilities to dollar equivalents are flawed. Even when computed reasonably, the results often seem to defy commonly held beliefs about prices and have limited strategic value for decision making.

Let us review a common technique for converting conjoint utilities to a monetary scale, and then we will suggest what we believe is a better approach. Here is how we can compute dollar equivalents from utilities. Imagine the following utilities for a single respondent for two attributes:

Attribute	*Utility*	*Price*	*Utility*
Feature X	2.0	$10	3.0
Feature Y	1.0	$15	1.0

For this respondent, a $5 change in price (from $15 to $10) reflects a utility difference of 2.0 (3.0 - 1.0). Therefore, every one utile change is equal to $2.50 in value (5 dollars/2.0 utiles). It then follows that feature X, being worth one utile more than feature Y, is also worth $2.50 more for this respondent.

We discourage the use of this type of analysis because it is a potentially misleading. Moreover, there is one practical problem that must be overcome if there are more than two price levels. Unless utility is linearly related to price, referencing different price points results in different measures of utiles per dollar. A common solution is to analyze the utility of price using a single coefficient. As long as the price relationship is approximately linear, this circumvents the issue.

Another problem arises when price coefficients are positive rather than negative as expected. This may happen for some respondents due to random noise in the data or respondents who are price insensitive. Such reversals would suggest willingness to pay more for less desirable features. One way to work around this is to compute dollar values of levels using average (across respondents) utilities, which rarely display reversals. Another approach to the problem is to ignore it,

assuming that the reversals are just due to random noise. A more proactive way to avoid reversals is to use an estimation method that enforces utility constraints, though there are potential drawbacks to this approach (Johnson 2000).

Additional complications arise when the price coefficient for a respondent is extremely small in absolute value, approaching zero. In that case, the dollar equivalents for incremental features become very large, approaching infinity. A typical way to handle this is to characterize the centers of the distributions using medians rather than means.

This type of analysis assumes that the conjoint method has accurately captured respondents' price sensitivity. Some conjoint methods (ACA and potentially any partial-profile method) tend to understate people's price sensitivity. This can result in inflated willingness to pay values.

But after taking the appropriate steps to compute reasonable dollar equivalents, the results are potentially misleading. Even when accurate price sensitivity has been estimated for each individual, an examination of average values will often reveal that respondents are willing to pay much more for one feature over another than is suggested by market prices. This often causes managers to disbelieve the results. However, we'll demonstrate later that such outcomes are to be expected when the monetary value of levels is computed in this way.

There are a number of fundamental problems with analysis based on average dollar values. First, it attempts to ascertain an average willingness to pay for the market as a whole. Firms usually offer products that appeal to specific targeted segments of the market. The firm is most interested in the willingness to pay among its current customers, or among buyers likely to switch to its products, rather than in an overall market average. Second, this approach does not reference any specific product, but instead considers an average product. We expect that a respondent's willingness to pay for an additional feature would depend upon the specific product that is being enhanced (e.g., a discount or a premium offering). Third, and most fundamental, this approach assumes no competition. Because a product purchase usually constitutes a choice among specific alternatives, the competitive context is a critical part of the purchase situation. To illustrate the fallacy of interpreting average dollar values, without respect to competitive offerings, consider the following illustration.

Economics on "Gilligan's Island"

Though perhaps loathe to admit it, many have watched the popular 1960s American TV program *Gilligan's Island*. The program revolved around an unlikely cast of characters who became marooned on an uncharted desert island. Each episode saw the promise of rescue. And when it seemed that the cast was finally going to make it off the island, the bumbling Gilligan always figured out some way to ruin the day.

One colorful pair of characters were the ultrarich Mr. Howell and his wife. Now, imagine that one day a seaworthy boat with capacity for two passengers pulls into the lagoon and offers passage back to civilization for a price to be negotiated. What is the dollar value of rescue versus remaining on the island for Mr. and Mrs. Howell? Mr. Howell might pull out his checkbook and offer the crew millions of dollars. Under the assumption of no competition, the dollar equivalent utility of rescue is astronomically high. However, it might be much lower for other islanders of more limited means, and the average dollar value for all of them would have little relevance to the captain of the boat in negotiating a price. What would matter is the dollar value of the potential customers and no one else.

Now, assume, just as Mr. Howell and the first crew are preparing to shake on the deal, a second, equally seaworthy ship pulls into the lagoon and offers its services for a fixed $5,000. Ever the businessman, Mr. Howell will choose the $5,000 passage to freedom.

What has happened here? Is the utility of getting off the island for Mr. Howell suddenly different? Has his price sensitivity changed? No. The amount Mr. Howell would be projected to pay under the assumption of no competition is indeed very different from the amount he will pay given the appearance of another boat.

If the first boat's crew had administered a conjoint interview to Mr. Howell and had computed his willingness to pay under the first method reviewed in this article, they would have concluded that he was willing to pay a lot more than $5,000. But how meaningful is that information in light of the realities of competition? The realistic problem for the boat captain is to figure out what price the market will bear, given the existence of competitive offerings.

We can illustrate this point using another example. What is your willingness to pay for a color monitor for your laptop computer versus a monochrome screen? Assume we conducted a conjoint analysis including monochrome versus color monitors. If we computed your willingness to pay for color over monochrome, we would likely find that the incremental value of color over monochrome is worth a thousand dollars or more. But how meaningful is this information to a laptop manufacturer given the fact that laptops with color monitors are readily available on the market at quite inexpensive prices?

Price Sensitivity Simulations in Competitive Context

For most marketing problems involving competition, the best strategic information results from carefully defined market simulations. If a firm wants to assess the incremental demand resulting from offering specific features for its product, or improving its degree of performance, it should be estimated within a realistic competitive context.

Estimates of market demand should also be based on specific objectives. For example, the objective may be to determine how much more may be charged for a product or service by offering a new feature without any net loss in market acceptance. This approach involves simulating a realistic competitive scenario with a conjoint market simulator. Assume four products (A through D) represent the current relevant products in the marketplace. Further assume that the firm is interested in offering an additional feature for product A, and wants to estimate what new price can be charged while maintaining the same share of preference. We first simulate a base case with products A through D placed in competition with one another, where A does not include the new feature. We record its share of preference (say, 15 percent). We then conduct another simulation in which we improve A by offering a new feature (and hold the competition B through D constant). The share of preference for A should increase (say, to 20 percent). We then perform additional simulations (again holding competition constant) raising the price of the new product A until its share of preference again drops to the original 15 percent. The difference in price between the more expensive improved Product A that captures 15 percent and the old Product A that captured 15 percent reflects the incremental monetary value that the market will bear for the new feature, given the competitive context and the objective of maintaining share constant.

Market simulations conducted using individual-level utilities are best for this analysis. Individuals have different preferences, and the company that produces product A is most concerned with retaining current product A customers and attracting new buyers among those most likely to switch. The company does not care so much about individuals who are extremely unlikely to buy its offerings. Market simulations based on individual utilities support such complex market behavior, focusing the willingness-to-pay analysis on a relevant reference product and critical individuals rather than the market whole. Such market simulations can also reveal complex competitive relationships between products, such as degree of substitution (cross-effects) and differences in consumer price sensitivity to each product.

In summary, the common practice of converting differences between attribute levels to a monetary scale is potentially misleading. The value of product enhancements can be better assessed through competitive market simulations. If the market simulations are conducted using individual utilities, such simulations focus the price/benefit analysis on the customers that are most likely to purchase the firm's product(s) rather than on an overall market average. They provide strategic information based on a meaningful context that enables better decisions, while avoiding the pitfalls of other ways of analyzing data. Of course, the success of the simulation approach hinges on a number of assumptions, including the following: (1) the conjoint method produces accurate measures of price sensitivity, (2) the relevant attributes have been included in the simulation model, and (3) the relevant competitive offerings are reflected in the simulation model.

Chapter 10

Market Simulators for Conjoint Analysis

The market simulator is usually considered the most important tool resulting from a conjoint analysis project. The simulator is used to convert raw conjoint (part-worth utility) data into something much more managerially useful: simulated market choices. Products can be introduced within a simulated market scenario and the simulator reports the percentage of respondents projected to choose each product. A market simulator lets an analyst or manager conduct what-if games to investigate issues such as new product design, product positioning, and pricing strategy. Market simulators are commercially available or can be constructed using spreadsheet programs.

10.1 What Is a Market Simulation?

A conjoint study leads to a set of utilities or part-worths that quantify respondents' preferences for each level of each attribute. These utilities can be analyzed in a number of ways. You can examine each respondent's utilities, but, if the number of respondents is large, this can be overwhelming. You might summarize the average utilities or compute average importances. You could create graphs and charts to display that information. But to many managers the results of conjoint analysis may seem abstract. Also, when we examine aggregate data or average responses, we may fail to detect important market segments—groups of consumers with unique and targetable preferences.

A good market simulator is like having all of your respondents gathered in one room for the sole purpose of voting on product concepts within competitive scenarios. The product concepts are defined in terms of the attributes and levels you used in the conjoint study. You walk into a virtual room, show them a market scenario (i.e., products A, B, and C), and they vote for the products they prefer. Millions of potential products and market situations could be evaluated, and your captive audience would never get tired, ask for lunch breaks, or require you to pay them by the hour.

How does a market simulator work? Let us suppose we were able to quantify how much people liked various flavors of ice cream. Let us refer to those preferences as utilities, and assume the following values for a given respondent:

Flavor	*Utility*	*Price*	*Utility*
Chocolate	0	$0.60	50
Vanilla	30	$0.80	25
Strawberry	40	$1.00	0

Using these utility values, we can predict how this respondent would choose between a vanilla cone for $0.80 or a strawberry cone for $1.00:

$$\text{Vanilla (30 utiles)} + \$0.80 \text{ (25 utiles)} = 55 \text{ utiles}$$

$$\text{Strawberry (40 utiles)} + \$1.00 \text{ (0 utiles)} = 40 \text{ utiles}$$

We predict that this respondent will prefer the vanilla cone.

Now suppose we had data for not just one, but 500 respondents. We could count the number of times each of the two types of cones was preferred, and compute a share of preference, also referred to as a share of choice. If 300 respondents are predicted to choose the vanilla cone for $0.80 and 200 respondents are predicted to choose the strawberry cone for $1.00, then we would obtain these shares of preference or choice:

Product Concept	*Share of Choice*
Vanilla at $0.80	$\frac{300}{500} = 0.60$
Strawberry at $1.00	$\frac{200}{500} = 0.40$

The simplest market simulation is a simulation that assumes a first-choice model. A first-choice model assumes respondents buy or choose the product alternative from the competitive set that has the highest total utility, as determined by summing the part-worth utilities associated with the levels describing each product. There are more sophisticated approaches for market simulations that are beyond the scope of this introductory chapter. These more advanced approaches include logit, Bradley-Terry-Luce, and randomized first-choice models.

10.2 Applications of Conjoint Simulations

Looking only at average preferences or part-worth utilities can mask important market forces caused by patterns of preference at the segment or individual level. Marketers are often not interested in averages, but in targetable segments or the idiosyncratic behavior of individuals.

Consider the following example with three respondents and their preferences or utilities for color:

Respondent	*Blue*	*Red*	*Yellow*
Manny	50	40	10
Moe	0	65	75
Jack	40	30	20
Average	30	45	35

Looking only at average utilities, we would pronounce that red is the most preferred color, followed by yellow. But if one of each color was offered to each respondent, red would never be chosen under the first-choice model, while yellow would be chosen once, and blue twice—the exact opposite of what aggregate utilities suggest. While this is a hypothetical example, it demonstrates that average utilities do not always tell the whole story. Many similar, complex effects can be discovered only through conducting simulations.

We can use simulators to answer basic questions about preference and shares of choice. We can use them to study the competitive environment and market segments. Furthermore, we can use the results of simulations to guide strategic decision making. Here are some of the benefits and applications of conjoint simulators:

- Conjoint simulations transform raw utility data into a managerially useful and appealing model: that of predicting market choice (share of preference) for different products. Under the proper conditions, shares of preference quite closely track with the idea of market share—something almost every marketer cares about.
- As demonstrated earlier, conjoint simulations can capture idiosyncratic preferences occurring at the individual or group level. These underlying effects can have a significant impact on preference for products in market scenarios. When multiple product offerings have been designed to appeal to unique segments of the market, capturing such effects is especially important for accurately predicting preference.
- Conjoint simulations can reveal differential substitutability (cannibalism or cross-elasticity effects) between different brands or product features. If two brands are valued highly by the same respondents (have correlated preferences), these brands will tend to compete more closely. Product enhancements by one of these brands will result in more relative share being lost by the correlated brand than by other, less similar brands within the same simulation. Examining aggregate utilities cannot reveal these important relationships.

- Conjoint simulations can reflect interaction effects between attributes. If the same respondents that strongly prefer the premium brand are also less price sensitive than those who are more likely to gravitate toward a discount brand, sensitivity simulations will reflect a lower price elasticity for the premium relative to the discount brand. A similar interaction effect can occur between many other types of attributes, such as model style and color.
- Conjoint simulators may be used to answer questions about new products and new product introductions. Given a current competitive environment, what product should I offer to maximize interest in my offering? How can I modify an existing product to capture more relative demand? A market simulator lets you input multiple products and place them in simulated competition with one another. Each product is defined using the attribute levels measured in the conjoint study (brands, colors, prices, speeds, warrantees, etc.). Therefore, if you have measured the relevant brands and features offered in the market, you can simulate a realistic market scenario within the market simulator. Within that market scenario, you can add a new product and see how well it competes. If the goal is to maximize share, offering the best features at the lowest price is often the trivial solution. The market simulator focuses on the demand side of the marketing equation; but it is also important to pay attention to the supply side and take the costs of producing different products/services into consideration. If you have cost information available to you, the market simulator permits you to investigate the incremental benefits of different features of a product relative to the cost of offering them.
- Conjoint simulators may be used to guide pricing strategy. What is the relative price sensitivity of different brands? If I raise my price by 10 percent, how will it affect my brand? How will it affect competitor's brands? You can conduct sensitivity analysis for attributes such as price using the market simulator to generate relative demand curves. The approach involves holding all other brands at a constant price and changing the price of a single brand, recording the relative share at each point for that brand along the price continuum.
- Conjoint studies can help us to answer questions about product bundles and product portfolios. What portfolio of products can I offer to appeal to different market segments and maximize overall share? If you have segmentation information (such as demographics or firmographics), you can investigate product formulations that appeal to different groups of respondents. It is likely that, by designing products that appeal uniquely to targetable segments, you can increase overall share for your product line or occupy a niche that is not currently being served.

The next four sections of this chapter provide more detailed examples of applications, focusing upon introducing new products, estimating demand curves and elasticities, designing products to appeal to market segments, and game theory to inform marketing strategy

For the next three sections you should assume the following three attributes, each with three levels:

Brand	*Style*	*Price*
A	X	$100
B	Y	$150
C	Z	$200

10.3 Introducing New Products

Let us assume that your company is interested in entering a market that currently consists of just two competitors. There are only three attributes that adequately describe the products and account for preference in the market: brand, style, and price. The two products are Mellow (Brand A, Style X, at $100) and Mild (Brand B, Style Y, at $200).

Your company has developed a new product called Middling that has Style Z. You think Middling may appeal to buyers, and you want to investigate its potential with respect to the two existing products. The first step, typically, is to simulate the existing market scenario. You use the market simulator to define the two existing products:

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200

Suppose a market simulation leads to the following shares of preference:

Product	*Share of Preference*
Mellow	64.3
Mild	35.7

In this simulation, we see that 64.3 percent of respondents preferred Mellow and 35.7 percent preferred Mild. Note that the buyers in the simulation are all assumed to choose a product, so the shares of preference across products in the simulation sum to 100 percent.

Let us assume that you have actual market share information about these two brands. You note that the shares reported above do not necessarily match the actual market shares. You accept this, however, recognizing that many factors

influence market shares in the real world that cannot be captured through conjoint analysis. You are principally interested in relative preferences, assuming that the marketplace is an equal playing field (equal distribution, awareness, effectiveness of sales force, and equilibrium long-range demand).

In the second stage of this simulation example, we'll define a new scenario that includes your company's proposed product: Middling (Brand C, Style Z, $150. You add another product to your simulation specifications:

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200
Middling	C	Z	$150

Running the simulation again might lead to the following shares of preference:

Product	*Share of Preference*
Mellow	42.5
Mild	21.3
Middling	36.2

You note that Mellow is still the most preferred product, but that your product Middling is preferred to Mild. Like any market research statistics computed from samples, shares of preference are not estimated without error. It is common to estimate a confidence interval to get a feeling for the degree of uncertainty due to sampling and measurement error associated with a given share of preference. Let us assume that the standard error reported for Middling in the simulation above was 1.53. The 95% confidence interval is computed by adding plus and minus 1.96 times the standard error to the estimated share of preference. In this example, the 95% confidence interval is 36.2 plus and minus $(1.96)(1.53) = 3.0$ share points, or the interval [33.2, 39.2].

You next may ask yourself what price you would need to charge to capture the same relative preference as Mellow. To simulate this, you lower the price slightly for your brand. Many simulators include the ability to interpolate between levels (straight line interpolation), so you can investigate even the smallest of price changes. As a first step, you decide to lower the price to $130 for Middling (while holding the specifications for Mellow and Mild constant). The new simulated shares are as follows:

Product	*Share of Preference*
Mellow	39.2
Mild	19.0
Middling	41.8

You have overshot the mark (Middling's share exceeds Mellow's share), so you try a slightly higher price than $130 and run the simulation again. You make repeated attempts until Middling's and Mellow's shares are equal. Let us assume that after a few more attempts, you discover that the price that makes your company's offering match the share of preference of the market leader is $136. Another way of thinking about this finding is that your proposed product Middling commands a $136 - $100 = $36 premium over Mellow. Respondents are indifferent between Brand A and Style X at $100 and Brand C and Style Z at $136.

10.4 Estimating Demand Curves and Elasticities

We will build upon the previous example during this section. We have computed shares of preference for three products that were defined using the following attribute level codes:

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200
Middling	C	Z	$150

The shares of preference for the products, as defined above, were as follows:

Product	*Share of Preference*
Mellow	42.5
Mild	21.3
Middling	36.2

Let us assume that we wanted to estimate a demand curve for your company's offering: Middling, in the context of the current competition and prices. We do this through sensitivity analysis. Recall that we measured three distinct levels of price: $100, $150, and $200. Note that we have already computed the share of preference for Middling when it is offered at $150 (36.2). To estimate the demand curve for Middling, we will need to conduct two additional simulations: a simulation with Middling at the lowest price ($100), and a simulation with Middling at the highest price ($200). For each of these simulations, we'll hold the Mellow and Mild product specifications constant.

To estimate Middling's share at the lowest price ($100), we use the following product specifications:

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200
Middling	C	Z	$100

After running another simulation, we may observe the following shares:

Product	*Share of Preference*
Mellow	33.9
Mild	15.6
Middling	50.5

We record Middling's share (50.5), and proceed to the next step. To estimate Middling's share at the highest price ($200), we use the following product specifications:

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200
Middling	C	Z	$200

We run the simulation again, and the following shares are reported:

Product	*Share of Preference*
Mellow	49.2
Mild	26.9
Middling	23.9

From these three separate simulation runs, we have the information we need to plot a demand curve for Middling, relative to the existing competitors and prices. Assuming that Mellow and Mild are held constant at current market prices, the relative shares of preference for Middling at each of the price points within the measured price range are as follows:

Middling Price	*Middling Share of Preference*
\$100	50.5
\$150	36.2
\$200	23.9

We have demonstrated how to estimate a demand curve for Middling, relative to the existing competitors at current market prices. If the goal is to estimate demand curves for all brands in the study, the usual procedure is to record the share for a brand at each price level while holding all other brands at the average or middle price. It is often interesting to plot these demand curves and look at the patterns of price sensitivity among brands and the different slope of the curves from one segment of the curve to the next. It is also common to want to characterize the degree of price elasticity using a single value, referred to as the price elasticity of demand:

$$\mathbf{E} = \frac{\textbf{percentage change in quantity demanded}}{\textbf{percentage change in price}}$$

If the brand or product follows the law of demand, as most products do, price increases lead to decreases in quantity demanded, and the elasticity is negative. The larger the absolute value of the elasticity, the more price sensitive the market is with respect to that brand or product.

Using the midpoints formula, we can compute the average price elasticity of demand across the demand curve for Middling:

$$\mathbf{E} = \frac{\frac{(q_2-q_1)}{(q_1+q_2)/2}}{\frac{(p_2-p_1)}{(p_1+p_2)/2}}$$

$$\mathbf{E} = \frac{\frac{(23.9-50.5)}{(50.5+23.9)/2}}{\frac{(200-100)}{(100+200)/2}} = \frac{-0.715}{0.667} = -1.073$$

Another way to compute the average price elasticity of demand (which can be more accurate if more than two price points along the curve have been estimated) is the log-log regression. One takes the natural log of prices and shares and regresses the log of share on the log of price (you can do this within a spreadsheet). The resulting beta is the average price elasticity of demand.

As with all conjoint simulation results, the resulting elasticities from conjoint simulators must be interpreted bearing in mind some assumptions. In particular, the degree of noise within the conjoint data is particularly relevant. For example, if the respondents to the conjoint survey answered in a more haphazard way compared to buyers in the real world, the price elasticities estimated from conjoint simulations may be uniformly understated (too insensitive). Even if this is the case, the relative price sensitivities for brands are still useful.

10.5 Designing Products for Market Segments

Customizing products to appeal to target segments or even individuals is a common theme in marketing. Many companies dedicate significant resources to developing a portfolio of products that it hopes will appeal to unique segments. For line extensions, the challenge for any company is to design new products that take share from its competitors without stealing an unacceptable amount of share from products within its existing line.

One common approach to designing an effective line extension is to use the conjoint data to segment the market into latent (not observed) market segments (sometimes referred to as clusters) that have similar preferences. These segments are called latent because they are not simply delineated based on an explicit variable such as gender, income, or company size. Rather, the underlying segments are revealed through a statistical segmentation technique such as cluster analysis or latent class modeling. Segments are formed with the goal of maximizing the differences in preference between groups while minimizing the differences in preference within groups. Once these latent segments have been identified, one can profile them in terms of other variables in the survey (i.e., demographics, usage, or media habits).

If you have enabled your market simulator to select respondents for analysis by segment, this can further enhance the power of the tool. For example, let's assume that a cluster analysis revealed three relatively different segments for the hypothetical example we've been using.

By examining the part-worths and importances for each group, you can gain insight into the product features that might appeal to each. You also should bear in mind the size of each segment, as this represents its demand potential. Consider the part-worth utility preferences in exhibit 10.1.

Attribute Level	Segment 1 (n = 128)	Segment 2 (n = 283)	Segment 3 (n = 216)
Brand A	39	-51	-44
Brand B	5	39	-29
Brand C	-44	12	73
Style X	61	-52	-34
Style Y	-23	45	-9
Style Z	-38	7	43
$100	56	55	50
$150	7	2	6
$200	-63	-57	-56

Exhibit 10.1. Part-worth utilities across segments

We can study the part-worths to learn about the differences among the segments. We can also use these preferences to simulate market choices for the market scenario we had used previously to obtain shares of preference across segments. Note that the shares below do not match the shares reported for earlier examples in this chapter. Since these results are for illustration only, no significance should be attached to this difference.

Product	*Brand*	*Style*	*Price*
Mellow	A	X	$100
Mild	B	Y	$200
Middling	C	Z	$150

	Shares of Preference			
Product	*Segment 1* ($n = 128$)	*Segment 2* ($n = 283$)	*Segment 3* ($n = 216$)	*Total* ($n = 627$)
Mellow	84.8	21.5	22.2	34.7
Mild	7.4	40.0	14.2	24.5
Middling	7.8	38.5	63.6	40.8

Let us assume your company produces Old Middling under Brand C with Style Z at $150. Your total share of preference is 40.8 percent. We see from the simulation by segment that yours is the most preferred product within segment 3, and the second-most preferred product in Segment 2. Mellow, the Brand A product, clearly dominates Segment 1, which is the smallest segment.

Let us assume that your company was interested in offering an additional product, call it New Middling. We could examine the table of part-worth preferences in exhibit 10.1 as a first step in formulating hypotheses about what additional product might be successful.

Starting in order, you may first consider Segment 1, but this segment does not seem to offer many opportunities for your brand. Brand A, offering Style X at a low price, has got this relatively small segment nearly wrapped up, and this segment does not seem very receptive to Brand C.

You next consider Segment 2, which seems to represent a better opportunity for your brand. It is a relatively large segment that prefers Mild under Brand B, but also seems receptive to the Brand C product, Old Middling. Note also that Segment 2 strongly prefers Style Y, but your company currently offers only Style Z. By offering a Style Y product, you might be able to convert some current Brand B customers from within Segment 2 to your product line.

You currently dominate Segment 3 and should probably not consider designing another product to appeal to this segment, since a good deal of the possible share to be gained from a new product would be taken from your existing product within that segment.

Let us simulate what happens if, in addition to your current product Old Middling (Brand C, Style Z, $150), you offer another product, New Middling (Brand C, Style Y, $200).

	Shares of Preference			
Product	*Segment 1* ($n = 128$)	*Segment 2* ($n = 283$)	*Segment 3* ($n = 216$)	*Total* ($n = 627$)
Mellow	82.2	17.2	18.6	31.0
Mild	7.2	32.0	11.9	20.0
Old Middling	6.8	27.7	47.8	30.4
New Middling	3.8	23.1	21.7	18.7

The new product has somewhat cannibalized the existing product, reducing its share from 40.8 (see the previous simulation) to 30.4, but has resulted in a relative overall gain of [(30.4 + 18.7)/40.8] - 1 = 20 percent in preference.

For line extension simulations you conduct, the answer will likely not be so clear and the process not so direct as we've shown here. You'd certainly want to investigate other product configurations to make sure you weren't overlooking even better opportunities to enhance share. You would also want to consider the cost implications of different options for line extensions. Also, you would probably want to conduct sensitivity analysis for the new product with respect to price, to determine a strategic price point (given your costs and market share goals).

Viewing the preferences and shares by segment is not required in designing an effective line extension. However, viewing the separate market segments can help you more quickly recognize patterns of preference, size the different segments of the market, and thus more easily arrive at a good solution.

This exercise of viewing segment-based preferences and designing products to fill heterogeneous needs is a useful approach. However, it would seem more efficient to let an automated search algorithm find an optimal product or set of products rather than to proceed manually. There are commercial software programs available that use different algorithms to find optimal or near-optimal solutions, even when the search space is extremely large. These optimizers use a variety of search algorithms, including exhaustive search, hill-climbing procedures, and genetic algorithms. Genetic algorithm and other search plug-ins for Excel are available, allowing researchers to construct their own simulators with optimization. More information on simulations and optimization approaches is available within the monograph by Krieger, Green, and Wind (2005).

10.6 Game Theory and Conjoint Analysis

Details of a conjoint model, such as specific part-worth utility estimates and importance scores, are often distracting to managers and key stakeholders and are typically distant from the actual decisions at hand. Executive stakeholders should instead focus on strategic product changes, possible competitive reactions, and the net benefit/loss that occurs under possible outcomes. This type of thinking reflects what is known in academics as game theory.

Game theory is a way to address strategic marketing decision-making in the face of uncertainty. If one is able to model possible actions (decisions about product formulation and price), identify potential competitive responses to those actions, and also assign metrics (such as market share, revenue, or profitability) to each outcome, then one can select the business action that will lead to the greatest likelihood of end-game success.

At the 2012 Sawtooth Software Conference, Chris Chapman (formerly of Microsoft and currently at Google) described a simple case study involving game theory and market simulations using conjoint analysis. The manufacturer of a PC accessory hardware device was considering the question of whether to add a feature X to its product line after learning that feature X was going to be available from component suppliers in the near future. This feature X component was analogous to a higher-speed processor in a computer. But feature X would add cost and might not make much difference in users' actual experience. Importantly, from a game theory modeling perspective, the product category had two dominant players, the manufacturing firm in question and a competitor. Feature X would be available to both players.

Various business stakeholders had differing opinions about whether feature X should be included. Some thought it would appeal to customers and grow category share, while others argued that it would simply add cost and make the category less profitable. Chapman and his co-author Love realized that this was an excellent opportunity to apply game theory. Additionally, the authors could model the possible outcomes of the game because they had fielded more than a dozen conjoint analysis studies in this product category and had the data needed to model a likely market outcome. Conjoint analysis had proven to be a robust and useful indicator of market outcomes in the category.

Chapman and Love modeled the situation as a two-player, simultaneous, one-step game with identical goals for the two players. Since each player had two options, to include feature X or not, the game had four possible outcomes:

(1) Neither firm provides X
(2) Firm provides X competitor does not
(3) Firm does not provide X, competitor does
(4) Both provide X

This section draws heavily on material presented at the 2012 Sawtooth Software Conference by Chris Chapman of Google and Edwin Love of Western Washington University.

	Estimated Market Share Outcome		
Strategies/Actions	**This Firm**	**Competitor**	**No Purchase**
Neither firm provides ***X***	23	44	33
Firm provides ***X***, competitor does not	61	20	19
Firm does not provide ***X***, competitor does	10	72	18
Both firms provide ***X***	29	54	17

*If the competitor provides **X**, then the best strategy for the firm is to provide **X** also and realize the hollow circle outcome, 29. If the competitor does not provide **X**, then the best strategy for the firm is to provide **X** and realize the shaded circle outcome, 61. This means that providing **X** is a dominant or winning strategy for the firm and the one to recommend to management whose goal is market share maximization.*

Exhibit 10.2. Game theory strategies and outcomes

Product executives identified the division's strategic goal as maximization of market share, specifically to gain share from the other player. This led Chapman and Love to compute the four sets of outcomes for each player as preference shares for likely product lines with and without feature X. They then computed the share for the players in each of the four outcome scenarios (as well as the shares expected to not purchase the product) using a comprehensive set of conjoint data. The expected shares for the possible outcomes are shown in exhibit 10.2.

Furthermore, executive management felt that it was quite likely that the competitor was going to include feature X. It can easily be seen that if the competitor offered X and the firm did not also offer X, the effect on the firm would be decidedly negative. However, if the firm provided X, the outcome in terms of share would be significantly improved regardless of the competitor's reaction.

Management of the firm was convinced by this analysis and ultimately included feature X in its product line. As it turned out, the competitor had not anticipated the firm's action and only belatedly added X to its product line. The competitor's sluggishness in introducing feature X was detrimental to its brand image.

Without the game theory model that convinced business stakeholders to introduce feature X, it is likely that neither the firm nor its competitor would have introduced feature X for one or more years. The firm would have missed out on an opportunity to advance its product line and to meet consumer demand. In a worst-case scenario, the firm would have risked incursion into the category by another brand. Instead, the category was improved by the firm, strengthening its position and delivering better, more highly desired products to consumers.

This illustration involved only four possible outcomes within a game. One often sees competitive games with many more potential outcomes, a large number of potential product changes, and many competitors. If the likelihood of various competitor reactions can be estimated, then expected payoffs can be calculated for each potential move by a firm.

10.7 Simulation Methods and Sample Sizes

Part-worth utilities can be used within a choice simulator to predict preference for different product concepts in competitive scenarios. There are various simulation methods, including the simple first-choice (maximum utility rule) and the logit or Bradley-Terry-Luce model. First-choice simulations assume that each respondent can choose or vote for only one product and that one alternative captures 100 percent of the share for each respondent. Shares of preference under the first-choice rule are proportions.

In contrast, logit or Bradley-Terry-Luce models let respondents choose products in a probabilistic manner. Suppose there are three products in a market scenario. Representing a respondent's preferences with a probabilistic model might show choice probabilities (0.6, 0.3, 0.1), but the first-choice rule would represent the probabilities as (1, 0, 0). The probabilistic model captures more information from each respondent and yields more stable share estimates. The standard errors for share predictions from logit or Bradley-Terry-Luce simulations are always smaller than under the first-choice rule. Therefore, if you plan to use the first-choice model, you will need larger sample sizes to stabilize share-of-choice estimates relative to probabilistic simulation models.

10.8 Interpreting the Output of Market Simulators

Under very controlled conditions (such as markets with equal information and distribution), market simulators often report results that closely match long-range equilibrium market shares. But conjoint utilities cannot account for many real-world factors that shape market shares, such as length of time on the market, distribution, out-of-stock conditions, advertising, effectiveness of sales force, and awareness. Conjoint analysis predictions also assume that all relevant attributes that influence share have been measured. Therefore, the share of preference predictions usually should not be interpreted as market shares, but as relative indications of preference.

Divorcing oneself from the idea that conjoint simulations predict market shares is one of the most important steps to getting value from a conjoint analysis study and the resulting simulator. While external-effect factors can be built into the simulation model to tune conjoint shares of preference to match market shares, we suggest avoiding this temptation if at all possible. No matter how carefully conjoint predictions are calibrated to the market, the researcher may one day be embarrassed by differences that remain. Also, using external effects often changes

the fundamental properties of the original simulation model, such as the price sensitivities and substitution rates among products (Orme and Johnson 2006).

10.9 Multi-Store Simulators

The assumption of equal distribution is often responsible for the greatest differences between actual market shares and simulated shares of preference. Fortunately, there is a correct and straightforward simulation method for this problem. A multi-store simulator provides an appropriate way to account for an unequal distribution of products across the market without changing the products' original price sensitivities or substitution rates (Orme and Johnson 2006).

A multi-store simulator allows the researcher to specify, in the simplest case, the percentage of the regions/stores that carry each product. Superior implementations specify which products are available within each region/store and how much volume each region/store accounts for. Respondents are then randomly selected (with probability proportional to store volume) to make simulated visits to multiple stores on each of hundreds or thousands of occasions and to make choices among available products. If the respondent locations are known, we assign respondents to visit the applicable regional stores, rather than using a random process of assigning respondents to stores. The multi-store simulator is not just a tool for adjusting simulated shares to reflect better the availability of products across the market (and, in turn, market shares), but it is also a tool that more directly accounts for substitution effects by recognizing which products compete directly with one another (because they tend to be offered within the same regions/stores).

Chapter 11

Maximum Difference Scaling

Maximum difference scaling (MaxDiff) has experienced a recent surge in popularity, especially among analysts familiar with conjoint and choice analysis. The percentage of conjoint software users employing MaxDiff has grown from 8 percent in 2005 to 54 percent in 2012 (Sawtooth Software 2012). Presentations featuring MaxDiff have won best paper awards at recent marketing science conferences, including ESOMAR, the Advanced Research Techniques (ART) Forum, and the Sawtooth Software Conference.

MaxDiff is used for measuring the importance or preference within a list of items, such as product features, brands, advertising claims, product packaging, and job-related factors. It is meant to replace standard (and problematic) ratings questions, in which we ask respondents to rate items on (typically) a five-point scale.

Although MaxDiff, also known as best-worst scaling, is not technically a conjoint method, it is very similar. MaxDiff is easier to understand than conjoint analysis, involves fewer pitfalls, and is applicable to a wide variety of problems. But MaxDiff is not a substitute for conjoint.

11.1 Motivation for Maximum Difference Scaling

As researchers, we are constantly measuring the importance of or preference for things. The default approach has been the rating scale, often formatted as a grid, such as the one in exhibit 11.1.

The good news about grid-style ratings is respondents can answer them very quickly, often in about five seconds or less per item. Thus, respondents can provide data on dozens of product features or brands, while leaving time for more questions that we like to pack into questionnaires. But is this really good news? Should speed alone be the goal?

A key problem with standard ratings of importance is that people say that most things are very important or extremely important. From a data analysis perspective, we would prefer them to use the full breadth of the scale in a discriminating way, without having so many items rated at the extremes. We would like to be

How important are these aspects of fast food restaurants to you?

	Not at all important	*Not very important*	*Somewhat important*	*Very important*	*Extremely important*
Has clean eating area	○	○	○	○	○
Has clean bathrooms	○	○	○	○	○
Some health foods items on the menu	○	○	○	○	○
You get your food quickly	○	○	○	○	○
Staff are dressed professionally	○	○	○	○	○
Prices are very reasonable	○	○	○	○	○
Your order is always filled right	○	○	○	○	○
Has a play area for children	○	○	○	○	○
Food tastes wonderful	○	○	○	○	○
Restaurant gives generously to charities	○	○	○	○	○

Exhibit 11.1. Typical ratings grid

able to discover differences in importance across items for each individual and differences across segments of respondents for each item. Unfortunately, we often see little variability because the standard rating scale allows respondents to give lazy, non-discriminating answers.

But the problem is even more insidious than we have described so far. While some respondents tend to favor the upper end of the rating scale (yea-saying), others respondents may gravitate toward lower scale points (nay-saying). Moreover, some conscientiously use the full breadth of the scale while others concentrate on a narrow section of the scale. These tendencies are referred to as scale use bias, and they harm our ability to get an accurate measurement of preferences. Scale use bias can be problematic when comparing groups of respondents to determine which group is most interested in a particular product feature. Is the observed difference real or just an artifact of the groups' scale use tendencies? Those involved in international research should be concerned if different cultures use rating scales differently. Indeed, comparing average ratings across countries can pose significant challenges.

Some researchers encourage respondents to provide a wider variety of responses both within and between items by adding points to the rating scale. But with more scale points, respondents often react by shifting their answers to the left or right, continuing to use few of the available scale points (such as the tendency to use every fifth or tenth point on a 100-point scale). Researchers have also tried to counter scale use bias by normalizing data across respondents—zero-centering the data and equalizing the variances. The resulting standardized scores are a bit awkward to present to decision-makers because they have positive and negative values. Furthermore, these post-processing steps do not resolve the fundamental problems with traditional rating scales.

When considering fast food restaurants,
among the four attributes shown here,
which is the most important and least important?

Most Important		Least Important
○	*Prices are very reasonable*	○
○	*Your order is always filled right*	○
○	*Food tastes wonderful*	○
○	*Some health foods items on the menu*	○

Exhibit 11.2. MaxDiff question

To add to the problem of scale use bias, some cultures or segments of the population have a difficult time communicating strength of preference using values on a rating scale. This is especially true for children and individuals with little education. MaxDiff avoids these issues. No scale is presented to respondents. There are no scale labels to misinterpret, so scale use bias is irrelevant. People of all ages, cultural backgrounds, and educational levels find it natural to make choices. MaxDiff also capitalizes upon the fact that people find it easier to identify extremes than to discriminate among items of middling importance. Making choices is common to the human experience; rating scales are not. MaxDiff resolves many problems with traditional rating scales.

11.2 Efficient Data Collection Mechanism

With MaxDiff, we begin with a list of items or attributes (typically eight or more) and the desire to measure them on a common scale of importance or preference. In each MaxDiff question, respondents are shown just a few of the items (typically four to six items at a time). For each set of items, respondents pick the most and least important (or most and least preferred), as shown in exhibit 11.2.

Respondents typically complete about eight to twenty-four such questions. Across the questions, each item is seen many times by each respondent. The questions are designed carefully so that each item appears approximately an equal number of times. Furthermore, each item appears approximately an equal number of times with every other item.

Jordan Louviere, the inventor of the technique, originally called the method best/worst and later referred to it as maximum difference measurement. The latter label refers to the idea that respondents selecting the best and worst items

are identifying the pair of items with the maximum difference in importance or preference among all possible pairings of items within the question.

MaxDiff captures a great deal of information per unit of respondent effort. Consider a MaxDiff questionnaire with four items per set (items A, B, C, and D). If the respondent says that A is best and D is worst, we can infer a number of relationships (inequalities) among those four items. We of course know that A (best) is preferred to D (worst). But, we also know that A is preferred to both B and C. Furthermore, we know that B and C are also preferred to D. The only unknown is how B compares to C. Thus, with just two clicks, the respondent has provided information regarding five of the six possible paired comparisons within the set:

$$A > B, A > C, A > D, B > D, C > D$$

It is easy to see why MaxDiff may be considered an efficient mechanism for obtaining paired comparison information. The method of paired comparisons has been a mainstay of preference research for well over fifty years. Recent research has shown that MaxDiff works better than the traditional method of paired comparisons for scaling multiple items (Cohen 2003).

11.3 Analysis of Maximum Difference Data

Two common ways to analyze MaxDiff data are counting analysis and score estimation. Score estimation is often done using techniques such as aggregate logit, latent class, and hierarchical Bayes (HB).

Counting Analysis: Best and Worst Percentages

If the questionnaire design is near-orthogonal (meaning each item is shown approximately an equal number of times with every other item in the study) one may compute the percentage of times respondents choose each item as best (most important) or worst (least important). These two measures should run essentially counter to one another—items chosen most often as best will be chosen least often as worst. Table 11.1 shows hypothetical percentage best and worst results for ten items dealing with aspects of fast-food restaurants.

Best-Worst Percentage Differences

A simple way to combine the information from best and worst judgments is to subtract the percentage of times an item is selected worst from the percentage of times it is chosen best. For example, if item j is selected best 40 percent of the time and selected worst 10 percent of the time, the combined score for item j is 40 minus 10, or 30 percent. The combined scores range from a maximum of 100 percent (for an item always chosen best) to a minimum of -100 percent (for an item always chosen worst). Scores greater than zero indicate that an item is more likely to be chosen best than worst, whereas negative scores indicate that an item is more likely to be chosen worst than best. One can display the resulting scores

Table 11.1. Best and worst percentages from counts

Restaurant Characteristic	*Percent Best*	*Percent Worst*
Clean eating area	38.6	11.0
Clean bathrooms	29.4	18.6
Health food items	19.9	25.1
Get food quickly	20.8	27.6
Dressed professionally	18.3	30.7
Reasonable prices	23.1	23.8
Get order right	24.6	22.5
Play area	19.5	27.9
Tastes wonderful	36.1	13.5
Gives to charities	19.0	30.9

using bar charts. Such bar charts may be used to compare across segments of the population. Hypothetical percentage difference scores for males versus females are shown in figure 11.1.

Estimated Scores

A more precise way to estimate item scores involves applying a statistical model, such as aggregate logit, latent class, or hierarchical Bayes estimation (HB). Deciding to apply a model rather than just counting choices has benefits beyond better precision. The latent class technique, for example, divides respondents into groups that have similar preferences (needs-based segmentation). Hierarchical Bayes analysis provides stable and accurate scores for each individual, whereas counting results for one individual would typically be a bit less precise (Orme 2009b).

As with counting analysis, model estimation can lead to separate scores or weights using information from best choices alone, worst choices alone, or composite scores from best and worst choices. Score estimation involves dummy coding similar to what we would use in conjoint analysis. And, as with conjoint analysis, estimated scores typically include both positive and negative values. Note that items with negative scores are not necessarily undesirable to respondents—negative scores mean that these items are less desirable than those with larger or positive scores. Table 11.2 shows estimated scores (based on an aggregate logit model) for the restaurant study.

Some researchers find it easier to rescale positive and negative scores on a zero-to-100 scale, as shown in table 11.3. (Note that we could have employed a similar rescaling with the combined counts in figure 11.1.)

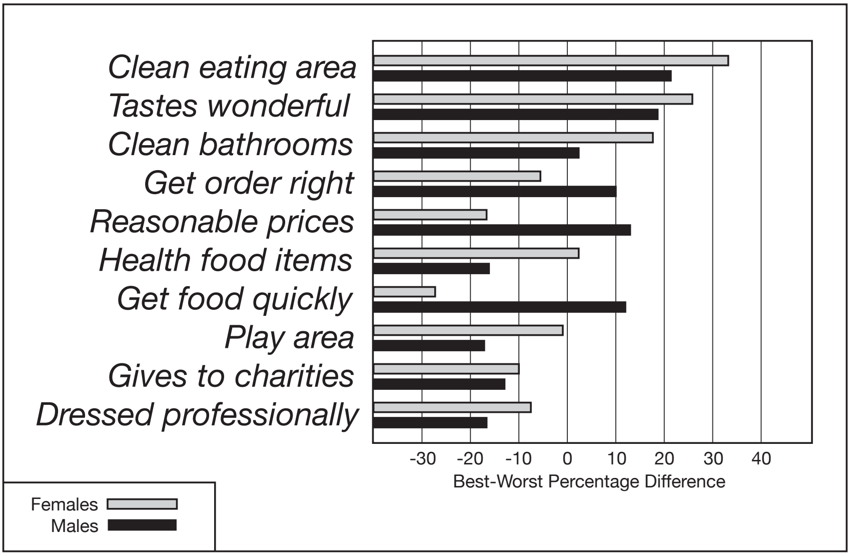


Figure 11.1. Combined importance scores by gender

Table 11.2. Scores from logit analysis (zero-centered scale)

Restaurant Characteristic	*Importance Score*
Clean eating area	1.30
Tastes wonderful	1.18
Clean bathrooms	0.80
Get order right	0.39
Reasonable prices	0.21
Health food items	-0.17
Get food quickly	-0.34
Play area	-0.56
Gives to charities	-1.32
Dressed professionally	-1.49

Table 11.3. Scores from logit analysis (zero-to-100 scale)

Restaurant Characteristic	*Importance Score*
Clean eating area	100
Tastes wonderful	96
Clean bathrooms	82
Get order right	67
Reasonable prices	61
Health food items	47
Get food quickly	41
Play area	33
Gives to charities	6
Dressed professionally	0

Table 11.4. Scores from logit analysis (100-sum scale)

Restaurant Characteristic	*Importance Score*
Clean eating area	25
Tastes wonderful	22
Clean bathrooms	15
Get order right	10
Reasonable prices	9
Health food items	6
Get food quickly	5
Play area	4
Gives to charities	2
Dressed professionally	2

Another approach is to place the weights on a ratio scale. With the ratio scale, all values are positive and an item with a score of 20 is twice as preferred as an item with a score of 10. We obtain ratio measures like this by taking the antilog of each item's zero-centered score and rescaling to a new score, so that the sum of the scores across all items is 100. In Microsoft Excel, the antilog formula is *=EXP(x)*, where x is the value to be transformed. After this transformation, all scale values are positive. It is common to rescale item scale values so their sum is 100, as shown in table 11.4. Other scale transformations may be used. For example, we could transform in a way that ensures that the average scale score across items is 100. Although ratio scales have strong properties, some researchers dislike the fact that the exponential transformation tends to stretch the scale so that a few best items receive large scores, while the worst items tend to group near zero.

Disaggregate Analysis

The illustrations in this chapter have involved aggregate (pooled) analysis. Most researchers analyze MaxDiff using disaggregate methods that estimate scores for segments using latent class analysis or scores for individuals using hierarchical Bayes models. In these cases, scores are developed separately for each segment or individual, with the scores averaged across segments or individuals to reflect the population.

An interesting outcome with disaggregate analysis that sometimes surprises researchers is that the summary importance scores for the population can modestly change in rank order depending on whether one averages across the raw individual-level scores from HB (or segment-based latent class analysis) or averages across the same scores transformed to the ratio scaling. The exponential

transformation to the ratio (probability) scale cannot change the order of preference of items within the segment or individual. But, when averaging exponentially rescaled scores across segments or individuals, the resulting summary rank-order of preference can slightly change. For example, an item that appears fourth most important for the sample under the raw logit-based scaling may change to third or fifth in importance under ratio scaling.

11.4 Maximum Difference Scaling Versus Conjoint Analysis

Although there are many similarities between MaxDiff scaling and conjoint analysis, each method has unique characteristics. The most noticeable difference is that MaxDiff does not require a structured organization of attributes, each with two or more levels. We simply construct a list of items or attributes for measurement on a common scale. After we obtain scale scores, all items may be directly compared.

As with conjoint analysis, items in MaxDiff questionnaires can be prohibited from appearing with other items. But MaxDiff prohibitions are much less detrimental to obtaining stable score estimates than with conjoint analysis. We recently used simulated respondent data to test MaxDiff score estimation for an unrealistically large number of prohibited combinations and were impressed by the method's robustness. Recognize, however, that the simple method of counting analysis becomes less accurate when prohibitions are used. Model estimation with the statistical methods mentioned earlier is usually required to obtain proper scores when one imposes prohibitions.

Although one can use conjoint-looking attribute lists within MaxDiff, the method does not formally support a market simulation capability. MaxDiff is a method for prioritizing a list of items, placing them on a common preference or importance scale. Its focus is on contrasting items rather than learning how combinations of the items taken together (conjoined, as with conjoint) affect buyer preference for product wholes. As a result, the concept of adding scores to predict the desirability of a product made up of multiple features is not supported. That said, the inventor of the technique, Jordan Louviere, has claimed some success with MaxDiff in projecting preference for product concepts by adding the scores for their features. He has also used MaxDiff with attribute lists from conjoint analysis, calling it "best-worst conjoint." This approach to conjoint analysis has not gained traction or acceptance in the research community.

If the goal of the research is to understand and predict preference for product or service concepts made up of combined attributes, then conjoint analysis is appropriate. If the emphasis is on prioritizing a list of items or features, placing them on a common scale, then MaxDiff is preferred.

11.5 Concerns about Maximum Difference Scaling

A flexible and useful scaling technique, MaxDiff is not without its problems. Researchers note the following concerns:

- *MaxDiff surveys are long.* MaxDiff surveys are much longer than surveys using traditional ratings. The typical MaxDiff survey takes about three times as long to complete as a traditional ratings survey.
- *The information from "bests" may be different from "worsts."* Flipping the scores derived solely from "worst" judgments (by multiplying by -1) and comparing to those derived solely from "bests" can reveal statistically significant differences. That is not to say that the scores are substantially different from a managerial perspective or that they usually would lead to different recommendations. Generally, this lack of symmetry has been more of a concern to academics than to practitioners. A solution would be to drop the "worst" question from the MaxDiff questionnaire. But, to get the same amount of information, a best-only questionnaire would need to be much longer than a best-worst or MaxDiff questionnaire.
- *MaxDiff focuses equally on achieving stable estimates of both best and worst items.* Typically, managers are more concerned with discriminating among the top few (best) attributes. However, MaxDiff questionnaires focus equal attention on items at the extremes (both best and worst). Newer adaptive approaches to MaxDiff have been proposed that focus on obtaining stronger estimates of best items, while sacrificing a modest amount of precision for items near the bottom of the scale (Orme 2006).
- *There is controversy about MaxDiff's conforming to strict error theory.* Experts in statistics have argued that MaxDiff data do not formally conform to error theory consistent with logit analysis. Academics have argued this point, but practitioners have paid little attention. Logit scores are very consistent with respondent preferences observed via counting analysis, and, for practical purposes, there does not seem to be a problem in applying logit theory to MaxDiff. Researchers concerned about this technical point can obtain many of the benefits of MaxDiff by analyzing only the "bests" half of the MaxDiff questionnaire.
- *MaxDiff has an arbitrary scale origin.* The comparative nature of selecting a best and a worst item is a strength of MaxDiff, but it leads to a weakness—the scores have an arbitrary origin.

  Suppose we ask two respondents to evaluate the importance of eight characteristics of airline travel. Of course, there are many more aspects than eight to describe the travel experience, but we have chosen to study just eight. For the first respondent, assume we have chosen the eight characteristics that he or she finds most important. For the second respondent, there may be other characteristics, many more important than the eight we have chosen to measure. That is, for these two respondents, the eight character-

istics we have chosen differ in importance in an absolute sense. But there is no way to determine from their answers to MaxDiff questions that the first respondent values the eight characteristics more highly than does the second respondent.

With MaxDiff, we never ask respondents to indicate their feelings about the absolute importance of attributes. Rather, the data are relative in nature, and the scores reflect an arbitrary origin. Often, MaxDiff scale scores are zero-centered. It is also common to constrain a particular item score to be zero, with the other items scaled with respect to that zero score. Furthermore, if we transform to a zero-to-100 scale, we do not resolve the issue of an arbitrary scale origin. If we employ additional calibration or anchoring questions, we can attempt to give MaxDiff scores an absolute as well as a relative meaning. See Bacon et al. (2007), Orme (2009a), Lattery (2010), and Horne, Rayner, Baker, and Lenart (2012) for examples of these approaches. Although there is some allure to anchored scaling, it is not without its drawbacks, as discussed in the next section.

11.6 Anchored Scaling for MaxDiff

Researchers have recently been investigating ways to overcome MaxDiff's relative scoring (arbitrary scale origin) issue. For the purposes of interpretation, it could be helpful if a zero-anchoring score marked the difference between important and not important items.

Two principal methods for anchoring MaxDiff scores have been debated and tested: an indirect method proposed by MaxDiff's inventor, Jourdan Louviere, called dual-response, and a direct method proposed by Kevin Lattery (Orme 2009a; Lattery 2010). After each MaxDiff set, the dual-response approach employs questions that ask respondents if all, some, or none of the items within the set are important (or would be "worth \$100 to purchase"). The direct approach displays all the items after the MaxDiff sets have been completed and asks respondents which items are important and which not.

Even though these anchored scaling methods for MaxDiff seem to be removing relativity from the MaxDiff scores and establishing a reference anchor, it is not clear that the anchoring is absolute or directly comparable across people. Furthermore, it is not clear that anchored MaxDiff leads either to better managerial decision-making or to better market segmentation analysis. One key problem is that the interpretation of the descriptor "important" is left up to the individual. The meaning of "important" can differ from one individual to the next, depending upon cultural background. Research by Horne, Rayner, Baker, and Lenart (2012) shows that respondent country of origin can have a significant effect upon the position of the importance anchor.

Using either the indirect or direct methods of anchoring, MaxDiff scales allow the 800-pound gorilla of yea-saying and nay-saying bias (a key element of scale-use bias) to reenter the room. One of the greatest benefits of using standard

MaxDiff scaling is its ability to avoid scale-use bias and allow more direct comparison of cross-cultural results. So, the fact that anchoring methods for MaxDiff reintroduce aspects of scale-use bias is discouraging.

A simpler way of incorporating absolute anchoring points for MaxDiff (a way that may work well in some cases) is to use an item or items within the MaxDiff study that have direct monetary meaning. For example, in studying job-related conditions and benefits, it might be useful to include a level that says: "Receive an immediate $500 bonus." The scores for other levels may then be compared to this monetary-based reference level. Also note that it is possible to make the monetary reference point multi-leveled, such as "Receive an immediate $500 bonus," "Receive an immediate $750 bonus," and "Receive an immediate $1,000 bonus." This provides multiple monetary-grounded reference points for comparing the scores of other items in the study.

In other studies, such as those studying potential modifications to existing products, services, or (to be more specific) employment conditions, it might make sense to include a reference level reflecting "no change." That is, include an item like "No change to current work environment." That way, researchers can identify item scores that might have a negative effect relative to "no change."

11.7 Predictive Validity

Standard MaxDiff is more discriminating than standard ratings-based methods and, most importantly, is free from scale use bias. But enhanced discrimination is not the ultimate goal of item scaling. Rather, validity is paramount in terms of understanding buyers' true motivations, preferences, and, ultimately, in predicting their behavior.

Recent research comparing MaxDiff to standard ratings scales shows that MaxDiff is more accurate in predicting holdout choices and brand preferences than traditional rating scales (Cohen 2003; Chrzan and Golovashkina 2007). Because it is a relative newcomer, the evidence of MaxDiff's predictive validity is not as extensive as the evidence for conjoint methods' validity. MaxDiff has many elements in common with conjoint methods and has generated interest within the conjoint community. We expect further evidence of its usefulness and validity to be presented in upcoming conferences and publications.

Chapter 12

Adaptive Choice-Based Conjoint

Choice-Based Conjoint (CBC) is the most widely used conjoint technique today. The marketing research community has adopted CBC enthusiastically for several reasons. Choice tasks mimic what actual buyers do more closely than the ranking or rating tasks of traditional full-profile conjoint analysis. Choice tasks seem easy for respondents—everyone can make choices. And, equally important, there are solid statistical models for deriving part-worth utility estimates from choices.

Unfortunately, CBC is less informative than tasks involving ranking or rating of product concepts. Before making a choice, the respondent must examine the characteristics of, typically, three to six product concepts in a choice set, each described on multiple attributes. Yet, his or her choice reveals only which product is preferred, and nothing about strength of preference or the relative ordering of the non-preferred concepts. Moreover, CBC questionnaires are often viewed as dull and unengaging.

Aware of the limitations of standard CBC, Johnson and Orme (2007) of Sawtooth Software introduced a new conjoint method—adaptive choice-based conjoint (ACBC). The method is having a significant impact upon conjoint practice. In 2012, 13 percent of studies using Sawtooth Software conjoint methods employed ACBC (Sawtooth Software 2012).

12.1 Origins of Adaptive Choice-Based Conjoint

Despite its popularity as a method of marketing research, standard CBC has many limitations:

- Product concepts in many conjoint surveys are not close to the respondent's ideal. This can create the perception that the interview is not focused or relevant to the respondent.
- Respondents (especially in Internet panels) do choice tasks very quickly. We have observed that, after respondents warm up to a CBC task, they typically spend no more than twelve to fifteen seconds per choice set.

- To estimate part-worths at the individual level, we require each respondent to make many choices. But when a dozen or more similar choice tasks are given to respondents, the survey experience is seen as repetitive and boring. Many respondents are less engaged in the process than the researcher might wish.
- Suppose a respondent is keenly intent on a particular level of a critical attribute (a *must have* feature), but if there is only one such product available per choice task, then the respondent is left with two alternatives: select the product with the critical feature or select *none* (assuming the *none* option is available). Thus, for respondents intent on a few key levels of attributes, a CBC choice task featuring minimal level overlap does not encourage respondents to reveal fully their product preferences. We learn only about a few *must have* features. To complicate matters further, superficial answers lead to seemingly consistent data and good statistical fit. For this reason, tasks with at least some degree of level overlap (level repeating) are useful.

Many CBC respondents answer choice questions more quickly than would seem possible if they were giving thoughtful responses using an additive, compensatory model. Through the analysis of commercial CBC datasets, researchers have found that most respondent answers can be accounted for by simple screening rules involving a few attribute levels (Gilbride and Allenby 2004; Hauser et al. 2006; Johnson and Orme 2007). Combine this fact with the realization (by anyone who has answered a CBC questionnaire) that the experience seems repetitive and boring, and we are led to conclude that there is a need for a different way of conducting choice research.

Huber and Zwerina (1996) showed that choice designs are more statistically efficient when product alternatives within choice sets are nearly equal in utility. Their research gave rise to the term “utility balance.” But choice tasks with utility balance cannot be designed without knowledge of the respondent’s utilities, which is not available until after the interview. This chicken-and-egg problem has led to several attempts at adaptive methods, in which information from early choice tasks is used to create greater utility balance in later choice sets.

Early attempts at adaptive methods led to mixed results (Johnson et al. 2005). But these attempts relied upon the assumption that respondents use a simple additive/compensatory strategy, consistent with the logit rule. In the design of adaptive methods, we need to acknowledge the fact that respondents often ignore some attributes and rely on non-compensatory decision rules, such as screening based on a few *must-have* or *must-avoid* features.

How can we obtain better or more complete data from research participants, while preserving the simplicity of a choice task? Is it possible to devise a survey method that reduces respondent fatigue or boredom and acknowledges respondent use of non-compensatory rules, while providing sufficient information to estimate numerous attribute parameters?

12.2 A New Approach to Data Collection

Johnson and Orme (2007) proposed adaptive choice-based conjoint (ACBC) as a survey method to reflect more closely or mimic the purchase process and to encourage deeper thought processing from research participants. The goal was not to design choice tasks with the highest statistical efficiency, but rather to acquire better choice data.

Researchers have long recognized that buyers in high involvement categories deal with the complexity of choosing among potentially dozens or even hundreds of available products by first screening on key characteristics to develop a manageable consideration set. Then, to identify an overall winner within the consideration set, buyers typically weigh a wider variety of aspects as they evaluate the relative strengths and weaknesses of the considered products.

The Johnson and Orme (2007) approach recognizes that respondents employ screening rules. The approach asks respondents to make more detailed choices among product concepts that pass screening criteria. The aim is to encourage respondents to make choices more thoughtfully, much like they would in an actual purchasing situation.

The ACBC interview has several sections. Throughout the interview, there is an attempt to keep the respondent interested and engaged. The instructions appear on the screen in text, but as though they were spoken by a friendly interviewer. The image of an interviewer appears frequently at various places in the interview, from different perspectives and in different poses. The interviewer explains to the respondent that this is a simulation of a buying experience and gives a rationale for each interview section. See Sawtooth Software (2009) for an online ACBC interview about laptop computers.

Build-Your-Own Section

In the first section of the interview the respondent answers a build-your-own (BYO) questionnaire to introduce the attributes and levels, as well as to let the respondent design the product concept he or she would most likely purchase (given the attribute levels in the study). Past research has shown that respondents enjoy BYO questionnaires and answer them rapidly, and that the resulting choices have lower error levels than repetitive choices from standard CBC questionnaires (Johnson, Orme, and Pinnell 2006).

A screen for a BYO section of an interview about laptop computers is displayed in exhibit 12.1. Based on answers to the BYO questionnaire, the ACBC algorithm creates a pool of product concepts (around twenty-four to forty) that are near neighbors to the respondent's preferred product. Each concept in the pool is generated following a near-orthogonal design by altering a few of the attributes from the BYO-specified concept.

Please select the laptop computer you'd be most likely to purchase.
For each feature, select your preferred level.

Feature	Select Feature	Cost for Feature
Size:	○ 14 inch screen, 5 pounds (+ $750) ○ 15 inch screen, 6 pounds (+ $750) ○ 17 inch screen, 8 pounds (+ $1,000)	$
Brand:	○ Acer ○ Dell ○ Toshiba (+ $50) ○ HP (+ $100)	$
Processor:	○ Intel Core 2 Duo T5600 (1.86GHz) ○ Intel Core 2 Duo T7200 (2.00GHz) (+ $100) ○ Intel Core 2 Duo T7400 (2.16GHz) (+ $300) ○ Intel Core 2 Duo T7600 (2.33GHz) (+ $550)	$
Operating System:	○ Vista Home Basic ○ Vista Home Premium (+ $50) ○ Vista Ultimate (+ $150)	$
Memory:	○ 512 MB ○ 1 GB (+ $100) ○ 2 GB (+ $250) ○ 4 GB (+ $400)	$
Hard Drive:	○ 80 GB ○ 100 GB (+ $50) ○ 120 GB (+ $100) ○ 160 GB (+ $150)	$
Video Card:	○ Integrated video, shares computer memory ○ 128MB Video card, adequate for most use (+ $50) ○ 256MB Video card for high-speed gaming (+ $200)	$
Battery:	○ 3 hour ○ 4 hour (+ $100) ○ 6 hour (+ $200)	$
Productivity Software:	○ Microsoft Works ○ Microsoft Office Basic (Word, Excel, Outlook) (+ $150) ○ Microsoft Office Small Business (Basic + PowerPoint, Publisher) (+ $250) ○ Microsoft Office Professional (Small Bus + Access database) (+ $300)	$
	Total	$

Next

Source: Adapted from Sawtooth Software (2009).

Exhibit 12.1. ACBC build-your-own questions

Screening Section

In the second section of the interview, the respondent answers screening questions in which product concepts (substantially like the one the respondent configured in the BYO stage) are shown a few at a time. Here, the respondent is not asked to make final choices, but rather to indicate whether each product concept is a possibility. Exhibit 12.2 displays a portion of the screening section for the laptop study. The exhibit shows only three product concepts, whereas an actual screening task might show four or five concepts on one screen.

The screening section identifies *must-haves* and *must-avoids* (unacceptables). After each group of concepts has been presented, previous answers are scanned to see if there is any evidence that the respondent is using non-compensatory screening rules. For example, we might notice that a respondent has expressed interest in only one level of an attribute, in which case the individual is asked whether that level is an absolute requirement (a *must-have*).

Past research with ACA has suggested that respondents are quick to mark many levels as unacceptable that are probably just undesirable. To reduce this possibility, ACBC offers only cutoff rules consistent with the respondent's previous choices and allows the respondent to select only one cutoff rule per screen. After each new screen of products has been evaluated, the respondent has an opportunity to add an additional cutoff rule. After a screening rule is confirmed by the respondent, any products in the pool that have not yet been evaluated but that fail to meet the cutoff criterion are automatically marked as non-possibilities and eliminated from further consideration.

Choice Tasks Section

In the third section of the interview the respondent is shown a series of choice tasks that present the surviving product concepts (those marked as possibilities). These are presented in groups of three (triples), as shown in exhibit 12.3. At this point in the survey respondents should be evaluating concepts that are close to their BYO-specified product, concepts that they consider as possibilities and that strictly conform to any cutoff, *must-have*, or *must-avoid* rules. To facilitate respondent information processing, attributes that are tied (have common attribute levels across the concepts) are grayed out, allowing respondents to focus on attribute differences across the concepts. Tied attributes are typically the most important factors for the respondent (based upon already established cutoff rules).

In the choice tasks section, then, the respondent is encouraged to discriminate further among products based upon features of secondary importance. Note that the approach of graying out tied attributes captures the benefits of partial-profile conjoint methods (only varying a subset of the attributes) but within the more realistic full-profile context.